ÜRETKEN YAPAY ZEKÂ KULLANIMI İLE ÖNERİ SİSTEMİ GELİŞTİRİLMESİ

Ozan İŞİN

T.C.
BURSA ULUDAĞ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ÜRETKEN YAPAY ZEKÂ KULLANIMI İLE ÖNERİ SİSTEMİ GELİŞTİRİLMESİ

Ozan İŞİN
0000-0002-0476-4636

Prof. Dr. Tülin İNKAYA
(Danışman)

YÜKSEK LİSANS
ENDÜSTRİ MÜHENDİSLİĞİ ANA BİLİM DALI

BURSA – 2025
Her Hakkı Saklıdır

ÖZET

Yüksek Lisans Tezi

ÜRETKEN YAPAY ZEKÂ KULLANIMI İLE ÖNERİ SİSTEMİ GELİŞTİRİLMESİ

Ozan İŞİN

Bursa Uludağ Üniversitesi
Fen Bilimleri Enstitüsü
Endüstri Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Tülin İNKAYA

Son yıllarda, yapay zekâ tabanlı Büyük Dil Modelleri (Large Language Models, LLM) eğitim, sağlık, finansal hizmetler ve eğlence gibi birçok sektörde doğal dil işleme, karar destek sistemleri ve kişiselleştirilmiş önerilerin oluşturulmasında önemli roller üstlenmeye başlamıştır. Bu çalışmada, içerik tabanlı bir yemek tarifi öneri sistemi geliştirilmiş ve performansının iyileştirilmesinde LLM kullanımının etkileri incelenmiştir. Uygulamada LLM olarak OpenAI firmasına ait ChatGPT'den yararlanılarak, kullanıcılar tarafından seçilen malzemeler ve yemek tiplerine dayalı kişiselleştirilmiş tarif önerileri oluşturulmuştur. Arayüz üzerinden elde edilen veriler, komut oluşturma fonksiyonu tarafından işlenerek LLM'e aktarılmakta; modelden alınan yanıtlar ise kullanıcı değerlendirmeleriyle birlikte veri tabanında kaydedilmektedir. Farklı komut stratejilerinin (sıfır atışlı, az atışlı ve rol tabanlı) sistem performansı üzerindeki etkisi, doğruluk, işlem süresi ve maliyet gibi metrikler üzerinden değerlendirilmiştir. Bulgular, üretken yapay zekânın (ÜYZ) bağlamsal anlayış ve kişiselleştirme kabiliyetleri sayesinde, öneri sistemlerinin doğruluk ve hız gibi kritik ölçütlerinde önemli iyileşmeler elde edilebileceğini göstermektedir. Aynı zamanda, kullanıcı etkileşimi verilerinin dinamik bir şekilde işlenmesi yoluyla farklı malzeme tercihlerine uyum sağlanması, tarif önerilerinin kalitesini artırmaktadır. Bununla birlikte, ÜYZ kullanımının maliyet ve modelin kapalı/kara-kutu niteliği gibi bazı sınırlamaları bulunduğu da gözlemlenmiştir. Bu sınırlamaların gelecekte farklı veri setleri, kullanıcı profilleri ve alternatif LLM sürümleriyle araştırılması önerilmektedir.

Anahtar Kelimeler: Öneri sistemleri, üretken yapay zekâ, komut mühendisliği, büyük dil modelleri, yemek tarifi
2025, xv +98 sayfa.

ABSTRACT

MSc Thesis

DEVELOPMENT OF A RECOMMENDER SYSTEM USING GENERATIVE ARTIFICIAL INTELLIGENCE

Ozan İŞİN

Bursa Uludağ University
Graduate School of Natural and Applied Sciences
Department of Industrial Engineering

Supervisor: Prof. Dr. Tülin İNKAYA

In recent years, artificial intelligence-based Large Language Models (LLMs) have begun to play pivotal roles in natural language processing, decision support systems, and the generation of personalized recommendations across multiple sectors, including education, healthcare, financial services, and entertainment. In this study, a content-based recipe recommendation system was developed, and the impact of utilizing LLMs on improving its performance was examined. In the implementation, ChatGPT—an LLM developed by OpenAI—was employed to generate personalized recipe suggestions based on the ingredients and types of dishes selected by the users. Data gathered through the interface is processed by a prompt engineering function and then passed to the LLM; responses from the model are stored in a database along with user evaluations. The influence of various prompt strategies (zero-shot, few-shot, and role-based) on system performance was evaluated using metrics such as accuracy, processing time, and cost. The findings indicate that, thanks to the contextual understanding and personalization capabilities of generative artificial intelligence (GAI), recommendation systems can achieve significant improvements in critical metrics like accuracy and speed. Additionally, dynamically processing user interaction data to adapt to different ingredient preferences enhances the quality of the recipe suggestions. However, it has also been observed that the use of GAI entails certain limitations, such as cost and the model's opaque "black-box" nature. It is recommended that these limitations be investigated further using different datasets, user profiles, and alternative LLM versions in future research.

Keywords: Recommender systems, generative artificial intelligence, prompt engineering, large language models, food recipe
2025, xv +98 pages.

TEŞEKKÜR

Bu çalışmayı gerçekleştirirken sürecin tamamı boyunca desteğini esirgemeyen saygıdeğer tez danışmanım Prof. Dr. Tülin İNKAYA'ya, çalışmayı gerçekleştirmek için gerekli zamanı yaratabilmem için bana gerekli desteği ve esnek çalışma şartlarını sunan müdürlerim Serkan UZUN ve Gizem AKBEY'e, hayatım boyunca devamlı olarak sonsuz desteğini hissettiğim annem Vasfiye YÖNDEM'e teşekkürlerimi sunarım.

Ozan İŞİN
28/03/2025

İÇİNDEKİLER

Sayfa

ÖZET viii
ABSTRACT ix
TEŞEKKÜR x
SİMGELER ve KISALTMALAR DİZİNİ xii
ŞEKİLLER DİZİNİ xiii
ÇİZELGELER DİZİNİ xiv
1. GİRİŞ 1
2. KAYNAK ARAŞTIRMASI 3
3. MATERYAL ve YÖNTEM 8
3.1. Öneri Sistemleri 8
3.2. Büyük Dil Modeli 10
3.3. Komut Mühendisliği 11
3.4. Öneri Sistemlerindeki Komut Stratejileri 14
3.5. Yemek Tarifi Öneri Sistemleri 15
3.6. Komut Oluşturma Algoritması 16
3.7. Önerilen Yaklaşım 17
3.8. Önerilen Yaklaşımın Performans Değerlendirmesi 19
3.9. Performans Değerlendirme Ölçütleri 25
4. BULGULAR ve TARTIŞMA 29
4.1. Veri Seti 29
4.2. Uygulama Koşulları 29
4.3. Deneysel Tasarım 33
4.4. Deneysel Tasarımın Sonuçları 35
4.5. İstatistiksel Analiz 48
4.5.1. Kruskal-Wallis ve Epsilon Kare Testleri 49
4.5.2. Permütasyonel Varyans Analizi 54
4.6. Tartışma 61
5. SONUÇ 63
KAYNAKLAR 65
EKLER 69
EK 1 NDCG etkileşim grafikleri 69
EK 2 RMSE etkileşim grafikleri 70
EK 3 MAE etkileşim grafikleri 71
EK 4 Yanıt süresi etkileşim grafikleri 72
EK 5 Toplam jeton adedi etkileşim grafikleri 73
EK 6 Shapiro–Wilk testi sonuçları 74
EK 7 Anderson–Darling testi sonuçları 75
EK 8 Kovaryans matrisi homojenliği test sonuçları 76
EK 9 PERMANOVA sonuçları 80
EK 10 Deneysel sonuçlar 82
ÖZGEÇMİŞ 98

SİMGELER ve KISALTMALAR DİZİNİ

Kısaltmalar	Açıklama
ANOVA	Varyans Analizi (Analysis of Variance)
API	Uygulama Programlama Arayüzü (Application Programming Interface)
DCG	İndirgenmiş Kümülatif Kazanç (Discounted Cumulative Gain)
DNN	Derin Sinir Ağı (Deep Neural Network)
DOE	Deney Tasarımı (Design of Experiments)
GPT	Üretici Önceden Eğitilmiş Dönüştürücü (Generative Pre-trained Transformer)
HIN	Heterojen Bilgi Ağı (Heterogeneous Information Network)
iDCG	İdeal İndirgenmiş Kümülatif Kazanç (Ideal Discounted Cumulative Gain)
IGCN	Tümevarımsal Grafik Konvolüsyonel Ağ (Inductive Graph Convolutional Network)
iWDCG	İdeal İndirgenmiş Azaltılmış Kümülatif Kazanç (Ideal Weighted Discounted Cumulative Gain)
KTS	Kendall'ın Tau İstatistiği (Kendall's Tau Statistic)
LLM	Büyük Dil Modeli (Large Language Model)
LM	Dil Modeli (Language Model)
LoRA	Düşük Dereceli Uyumlama (Low-Rank Adaptation)
MAE	Ortalama Mutlak Hata (Mean Absolute Error)
PERMANOVA	Permütasyonel Varyans Analizi (Permutational Variance Analysis)
MANOVA	Çok Değişkenli Varyans Analizi (Multivariate Analysis of Variance)
NCL	Negatif Korelasyon Öğrenimi (Negative Correlation Learning)
NDCG	Normalleştirilmiş Azaltılmış Kümülatif Kazanç (Normalized Discounted Cumulative Gain)
NLP	Doğal Dil İşleme (Natural Language Processing)
ÖS	Öneri Sistemi (Recommender System)
PLM	Önceden Eğitilmiş Dil Modeli (Pre-trained Language Model)
RMSE	Hata Kareler Ortalamasının Kare Kökü (Root Mean Squared Error)
WDCG	Ağırlıklı İndirgenmiş Kümülatif Kazanç (Weighted Discounted Cumulative Gain)
WNDCG	Ağırlıklı Normalleştirilmiş İndirgenmiş Kümülatif Kazanç (Weighted Normalized Discounted Cumulative Gain)
ÜYZ	Üretken Yapay Zekâ (Generative Artificial Intelligence)

ŞEKİLLER DİZİNİ

Sayfa

Şekil 3.1.	Uygulama arayüzü	18
Şekil 3.2.	Önerilen yaklaşım akış şeması	20
Şekil 3.3.	Verilen algoritma ile oluşturulan komut	22
Şekil 3.4.	Verilen algoritma ile alınan yanıt	24
Şekil 3.5.	Önerilen yaklaşıma ait sistem modülleri, kullanıcı etkileşimleri ve LLM arasındaki akış	24
Şekil 4.1.	Deneysel akış şeması	32
Şekil 4.2.	Sistem modülleri ve LLM arasındaki akış şeması	33
Şekil 4.3.	WNDCG için ana etkiler grafikleri	36
Şekil 4.4.	NDCG için ana etkiler grafikleri	37
Şekil 4.5.	RMSE için ana etkiler grafikleri	39
Şekil 4.6.	MAE için ana etkiler grafikleri	41
Şekil 4.7.	Yanıt süresi için ana etkiler grafikleri	42
Şekil 4.8.	Harcanan jeton adedi için ana etkiler grafikleri	44
Şekil 4.9.	WNDCG için etkileşim grafikleri	48

ÇİZELGELER DİZİNİ

Sayfa

Çizelge 4.1. Uygulamaya ait parametreler 31
Çizelge 4.2. Kruskal-Wallis testi sonuçları 51
Çizelge 4.3. Epsilon Kare testi sonuçları 53

1. GİRİŞ

Günümüzde, dijital ortamlarda milyonlarca kullanıcının etkileşimde bulunduğu platformlar, kişiselleştirilmiş içerik sunma ihtiyacını artırmaktadır. Bu bağlamda, öneri sistemleri giderek daha fazla önem kazanmaktadır. Geleneksel öneri sistemlerinin sınırlılıklarını aşmak ve kullanıcıların ihtiyaçlarına daha etkili yanıtlar sunabilmek adına üretici yapay zekânın yeni ufuklar açabileceğine inanılmaktadır.

Öneri sistemleri, kullanıcıların geçmiş etkileşimlerini analiz ederek kişiselleştirilmiş öneriler sunmak için karmaşık algoritmalar kullanır. Kullanılan bu öneri sistemi algoritmaları birçok çalışmaya temel oluşturmuştur. Bu algoritmaların belirli konulara uygun şekilde eğitilmesi, sınırlı veri setleriyle çalışması ve yerel bilgisayarların teknik sınırlamaları gibi zorlukları aşmak için, karmaşık algoritmalar yerine mevcut üretken yapay zekâ (ÜYZ) teknikleri örneklerinden biri kullanılmaktadır. ÜYZ insan yapımı çıktılar üreten, denetimsiz veya kısmen denetimli bir makine öğrenimi çerçevesidir (Hu, 2023 ve Jovanović, 2023).

Bu tezde; üretken yapay zekânın öneri sistemleri alanındaki kullanım potansiyeli, mevcut kısıtları ve avantajları incelenmektedir. ÜYZ'nin insan yapımı çıktılar üretebilme özelliği, öneri sistemlerinde daha özgün öneriler sunmayı mümkün kılarken, geniş kapsamlı veri işleme yeteneği sayesinde de sınırlı veri seti veya teknik altyapı sorunlarının aşılmasına katkı sağlar. Böylelikle öneri sistemlerinin performansının yükseltilebileceği ve kullanıcılara daha etkili, kişiselleştirilmiş içerikler sunulabileceği öngörülmektedir.

Mevcut literatür, öneri sistemlerinin temelde karmaşık algoritmalardan yararlandığını, ancak bu algoritmaların veri çeşitliliği ve kalite gibi faktörlere duyarlı olduğunu ortaya koymaktadır. Bu noktada ÜYZ, veri eksikliklerini tahmin yoluyla doldurabilme, yeni ve özgün öneriler oluşturma gibi avantajlarıyla devreye girerek öneri sürecini zenginleştirebilir. Yine de bu teknolojinin uygulanmasında, modelin eğitimi için gereken yüksek hesaplama gücü veya etik hususlar gibi konular da göz önünde bulundurulması gereken sınırlılıklar arasındadır.

Bu tez, beş ana bölümden oluşmaktadır. Giriş bölümü, çalışmanın amacı ve üretken yapay zekânın öneri sistemlerindeki rolünü tanıtmaktadır. Kaynak Araştırması bölümünde, literatürdeki önceki çalışmalar ve mevcut bilgi boşlukları incelenmiştir. Materyal ve Yöntem bölümü, kullanılan veri seti ve öneri sisteminin geliştirilmesi için izlenen yöntemleri açıklamaktadır. Bulgular ve Tartışma bölümünde, yapılan testler ve elde edilen sonuçlar ve deneylerden elde edilen verilerin analizleri yer almaktadır. Sonuçlar bölümü, çalışmanın genel değerlendirmesini ve gelecekteki araştırmalara yönelik önerileri sunmaktadır.

2. KAYNAK ARAŞTIRMASI

Öneri sistemleri, kullanıcıların ihtiyaç ve tercihlerini kavramak amacıyla büyük veri analitiği ve makine öğrenimi tekniklerini kullanan önemli bir teknoloji olarak öne çıkmaktadır. E-ticaret, sağlık, medya ve eğitim gibi çeşitli alanlarda yaygın bir şekilde uygulanan bu sistemler, doğru içeriği doğru kullanıcıya ulaştırarak hem kullanıcı memnuniyetini artırmakta hem de işletmelere değer sağlamaktadır. Son yıllarda, büyük dil modelleri (Large Language Model, LLM) ve üretken yapay zekâ (ÜYZ) tekniklerindeki gelişmeler, öneri sistemlerinin performansını yükseltme ve kişiselleştirme kabiliyetlerini güçlendirme açısından yeni fırsatlar sunmaktadır. Bu teknolojiler, geleneksel yöntemlerin ötesine geçerek daha esnek, ölçeklenebilir ve yenilikçi yaklaşımlarla öneri sistemlerinde dönüşüm sağlamaktadır.

Li vd. (2024) çalışmasında, öneri sistemlerinde kullanılan yöntemlerin belirli senaryolardaki performansları ve etkileri ele alınmıştır. Araştırmada, artırımlı grafik evrişim ağı (Incremental Graph Convolutional Network, IGCN) soğuk başlangıç ve kullanıcı-öğe etkileşim eksikliği gibi problemleri gidermek amacıyla meta-öğrenme yaklaşımlarını başarıyla uygulamış, veri yoğun ortamlar için etkili çözümler sunmuştur. Benzer şekilde, komşuluk geliştirmeli karşıt öğrenme (Neighborhood-enhanced Contrastive Learning, NCL) teknikleri, kullanıcı ve öğe arasındaki komşuluk ilişkilerini güçlendirerek grup öneri sistemlerinde önemli bir başarı ortaya koymuştur. Ayrıca, heterojen bilgi ağı (Heterogeneous Information Network, HIN) kullanılarak kullanıcı ve öğe ilişkileri bağlamsal bilgilerle zenginleştirilmiş; bilgi grafiği tabanlı sistemler ise sağlık ve e-ticaret gibi alanlarda, yüksek yapılandırılmış verilerden yararlanarak kullanıcı tercihlerini daha iyi analiz etmiştir. Derin öğrenme temelli hibrit sistemler ise kullanıcı geçmişi, sosyal bağlam ve öğe içeriklerini bütünleştirerek doğruluk ve gerçek zamanlı performansı artırmış, karmaşık kullanıcı ihtiyaçlarını karşılama konusunda önemli katkılar sağlamıştır.

Kumar vd. (2024) çalışması, öneri sistemlerinin gelişen yöntemlerini, sınıflandırmalarını ve çeşitli bağlamlardaki performanslarını detaylandırmıştır. İçerik tabanlı filtreleme yöntemleri, kullanıcıların geçmiş tercihlerine dayalı öneriler sunarken soğuk başlangıç ve

veri seyrekliği gibi sorunlara çözüm üretmiştir. İşbirlikçi filtreleme yaklaşımları ise kullanıcı-öğe etkileşimlerini kullanarak, gruplar arası benzerlikleri değerlendirip özellikle model tabanlı işbirlikçi filtreleme, derin öğrenme ve makine öğrenimi teknikleriyle uzun ve kısa vadeli kullanıcı tercihlerinin bütünleştirilmesinde başarılı sonuçlar elde etmiştir. Derin sinir ağı (Deep Neural Network, DNN) tabanlı yöntemler, karmaşık kullanıcı-öğe ilişkilerini modelleme kapasitesini artırarak hem doğruluk hem de esneklik sağlamış; örneğin, sinir ağları oturum tabanlı önerilerde zaman dinamiklerini ve sıralı kullanıcı davranışlarını dikkate alırken, grafik sinir ağları ise etkileşimleri grafik temsilleriyle zenginleştirmiştir. Çalışma, ayrıca bağlama duyarlı, açıklanabilir ve adil öneri sistemleri gibi yeni yönelimlere de değinerek, bu sistemlerin kişiselleştirme ve kullanıcı güvenini artırma potansiyelini vurgulamaktadır.

Lanling vd. (2018) tarafından yapılan araştırmada, büyük dil modellerinin (LLM), özellikle ChatGPT'nin, genel görevlerde gösterdiği etkileyici başarılar ve öneri sistemlerindeki potansiyel uygulamaları ele alınmıştır. Bu çalışmada, LLM'nin öneri sistemleri olarak kullanımı değerlendirilmiş ve öneri görevlerinde etkin kullanımına yönelik bir çerçeve sunulmuştur. Erişilebilirlik, ayarlama stratejileri, model mimarisi, parametre ölçeği ve bağlam uzunluğu gibi faktörlerin öneri sonuçlarına etkisi incelenmiştir. Ayrıca, görev açıklamaları, kullanıcı ilgi modellemesi, aday öğelerin oluşturulması ve komut stratejilerinin etkileri de değerlendirilmiştir.

Xu vd. (2018) ise, büyük dil modellerinin (LLM'ler) öneri sistemlerinde kullanım potansiyelini kapsamlı bir şekilde analiz ederek literatürdeki önemli bir boşluğu doldurmayı hedeflemiştir. Araştırmada ProLLM4Rec adlı genel bir çerçeve geliştirilmiş; bu çerçeve aracılığıyla LLM'lerin farklı öneri sistemleri senaryolarında nasıl entegre edilebileceği sistematik olarak incelenmiştir. Çalışma, LLM'lerin öneri modeli olarak doğrudan kullanılması, mevcut öneri modellerinin iyileştirilmesi ve sanal kullanıcılar oluşturarak öneri simülatörü olarak görev yapması olmak üzere üç temel paradigmada ele alınmıştır. Ayrıca, görev tanımı, kullanıcı ilgi modellemesi, aday öğelerin oluşturulması ve yönlendirme stratejilerinin öneri performansı üzerindeki etkileri, detaylı deneylerle değerlendirilmiştir. MovieLens-1M ve Amazon Books veri setleri üzerinde yapılan geniş deneyler, LLM'lerin genel bilgi ile kullanıcı davranışlarını birleştirme yeteneği sayesinde

geleneksel öneri sistemlerine kıyasla belirgin avantajlar sunduğunu göstermiştir. Xu ve ekibi, bu çalışma ile LLM'lerin öneri sistemleri alanında hem teorik hem de uygulamalı anlamda katkılar sağlayarak gelecekteki araştırmalara sağlam bir temel oluşturmuştur.

Di Palma vd. (2023) tarafından yapılan çalışma, ChatGPT'nin tavsiye sistemleri bağlamındaki potansiyelini değerlendirmek amacıyla kapsamlı bir değerlendirme hattı önermektedir. Araştırmada, sıfır atış ve rol yapma temelli komut (prompt) stratejileri kullanılarak ChatGPT-3.5 ve ChatGPT-4 modellerinin, kitap, müzik ve film gibi farklı alanlarda Top-N tavsiye, soğuk başlangıç ve yeniden sıralama senaryoları üzerinden performansları incelenmiştir. Çalışma, geleneksel işbirlikçi filtreleme ve içerik tabanlı yöntemlerle karşılaştırıldığında, özellikle kitap alanında daha yüksek doğruluk ve kişiselleştirilmiş öneriler sunduğunu; ancak çeşitlilik ve popülerlik yanlılığı gibi alanlarda belirli sınırlamalara sahip olduğunu ortaya koymaktadır. Bu sonuçlar, ChatGPT tabanlı öneri sistemlerinin potansiyelini ve aynı zamanda üretken modellerin özgünlük, tekrarlama ve dış alan önerileri gibi sorunlarını ele alabilmek için geliştirilmiş kapsamlı bir değerlendirme metodolojisinin gerekliliğini vurgulamaktadır.

Liu vd. (2023)'nin çalışması, geleneksel tavsiye yöntemlerinin görev-spesifik yapılarına alternatif olarak, ChatGPT'nin geniş çaplı dil ve dünya bilgisini tavsiye görevlerine aktarma potansiyelini incelemektedir. Çalışmada, ChatGPT, ince ayar yapılmadan sıfır atış ve az atışlı komut stratejileri kullanılarak, puan tahmini, sıralı öneri, doğrudan öneri, açıklama üretimi ve inceleme özetleme olmak üzere beş farklı senaryoda değerlendirilmiştir. Amazon Beauty veri seti üzerinde gerçekleştirilen deneyler, modelin bazı görevlerde umut verici sonuçlar elde ettiğini, diğerlerinde ise erken dönem basit modellerin performans seviyelerine yakın çıktığını ortaya koymuştur. Ayrıca, yapılan insan değerlendirmeleri, ChatGPT'nin sağlanan bilgiyi anlama ve daha net, makul sonuçlar üretme konusundaki yetkinliğini destekler niteliktedir. Bu çalışma, ChatGPT'nin tavsiye sistemlerine entegrasyonunun mevcut sınırlamaları ve gelecekteki iyileştirme olanakları üzerine önemli çıkarımlar sunarak, alandaki ileri araştırmalara zemin hazırlamaktadır.

Peng vd. (2023)'nin çalışması ise, ön eğitimli dil modeli (Pre-trained Language Model, PLM) tabanlı öneri sistemlerinin potansiyelini ele almıştır. Bu araştırma, farklı eğitim paradigmalarıyla öğrenilen modellerin bilgi çıkarma ve aktarma yöntemlerini inceleyerek, öneri performansını artırmada genellik, seyreklik, verimlilik ve etkililik açılarından değerlendirme yapmıştır. Ayrıca, PLM tabanlı öneri sistemlerini farklı eğitim stratejileri ve hedefleri doğrultusunda sınıflandıran bir taksonomi sunulmuştur.

Fan vd. (2023) ise, e-ticaret ve web uygulamalarının yaygınlaşması bağlamında öneri sistemlerinin önemine dikkat çekmiş; derin sinir ağlarının (DNN) bu sistemlerin geliştirilmesindeki rolünü irdelemiştir. Bununla birlikte, DNN tabanlı yöntemlerin bazı sınırlamaları göz önünde bulundurulduğunda, büyük dil modellerinin (LLM) bu sınırlamaları aşmada kullanılabileceği vurgulanmıştır. Çalışma, LLM destekli öneri sistemlerinin önyükleme, ince ayar ve yönlendirme paradigmalarını kapsamlı bir şekilde ele almıştır.

Friedman vd. (2023) tarafından benimsenen yaklaşımda, konuşma tabanlı öneri sistemi (Conversational Recommendation System, KTS) öne çıkarılmış; büyük dil modellerinin (LLM) doğal konuşma yetenekleriyle entegre edilerek kullanıcı tercihlerini anlama, esnek diyalog yönetimi ve açıklanabilir öneriler sunma gibi yeni uygulamalar önerilmiştir. Ayrıca, eksik konuşma verisi durumlarında LLM tabanlı kullanıcı simülatörleri ile sentetik konuşmalar üretilebilmesi sağlanmıştır. Bu yaklaşımlar, RecLLM adlı büyük ölçekli bir KTS'nin YouTube videoları üzerinde uygulanmasıyla örneklendirilmiştir.

Harte vd. (2023) çalışması, ardışık öneri problemleri ve LLM'lerin bu sistemleri nasıl oluşturup geliştirebileceği üzerine odaklanmıştır. Araştırmada üç farklı yaklaşım geliştirilmiş; BERT tabanlı öneri modeli (BERT4Rec) ile büyük dil modeli (LLM) gömülü yapısının birleşiminin performansı %15-20 oranında artırdığı gözlemlenmiştir. Ayrıca, basit bir LLM gömülü yapı kullanarak anlamca ilişkili öğelerin vurgulanmasının rekabetçi sonuçlar üretebileceği tespit edilmiştir.

Dai vd. (2023) çalışması, ChatGPT'nin öneri sistemlerindeki kullanımını incelemiştir. Araştırmada, ChatGPT'nin çeşitli doğal dil işleme (Natural Language Processing, NLP)

görevlerinde elde ettiği önemli başarıların yanı sıra, öneri sistemleri alanındaki yeteneklerinin tam olarak ortaya konulamadığı belirtilmiştir. ChatGPT'nin öneri kabiliyetlerini geliştirmek amacıyla, belirli alanlara yönelik üç farklı öneri politikası formatına odaklanılmış; deneyler ise ChatGPT'nin liste bazlı sıralamada maliyet ve performans arasında dengeli bir çözüm sunduğunu göstermiştir. Bu çalışma, ChatGPT'nin öneri görevlerine uyum sağlama potansiyelini ortaya koymaktadır.

Bu tezin özgün ve yenilikçi yönü, geleneksel karmaşık algoritmalar yerine üretken yapay zekâ (ÜYZ) teknolojilerinden yararlanarak öneri sistemleri geliştirmeye odaklanmasıdır. Araştırma, büyük dil modellerinin (LLM) farklı yönlendirme stratejileri kullanılarak kullanıcı niyetlerini daha iyi kavramasını sağlayarak, daha kişiselleştirilmiş ve doğru öneriler sunmayı amaçlamaktadır. Ayrıca, bu çalışma; öneri sistemlerinde nadiren karşılaşılan, yemek tariflerine dayalı bir uygulama örneği sunmasıyla, insanların günlük yemek hazırlama alışkanlıklarını değiştirme, onlara kolaylık sağlama ve yaratıcı seçenekler sunma potansiyeli taşımaktadır. Bunun yanında, yerel bilgisayarların teknik kısıtlamaları ve sınırlı veri setleri gibi zorlukları aşmak amacıyla, ÜYZ'nin esnekliği ve ölçeklenebilirliği kullanılarak öneri sisteminin performansının iyileştirilmesi hedeflenmiştir. Bu yaklaşım, öneri sistemlerinin etkinliğini artırmak ve yeni kullanım alanları oluşturmak için yenilikçi bir yöntem sunmaktadır.

3. MATERYAL ve YÖNTEM

Bu bölümde, çalışmada kullanılan materyal ve yöntem açıklanmıştır.

3.1. Öneri Sistemleri

Öneri sistemleri, kullanıcılara özelleştirilmiş öneriler sunarak ürün seçimi, müzik tercihi veya haber tüketimi gibi karar alma süreçlerine destek sağlayan yazılım araçları ve teknikleridir (Ricci vd., 2010). "Öğe" terimi, bu sistemlerin sunduğu önerileri tanımlar ve genellikle belirli bir kategoriye (örneğin, CD'ler, haber makaleleri) özgüdür. ÖS'nin tasarımı, kullanıcı arayüzü ve temel öneri algoritmaları, hedeflenen alanda etkili ve uygun öneriler sunmak üzere özelleştirilmiştir. Bu sistemler, özellikle çeşitli seçenekler arasında doğru seçimi yapmak için yeterli deneyimi veya uzmanlığı olmayan bireyleri hedef almaktadır (Ricci vd., 2010).

Öneri sistemleri; Amazon.com, YouTube, Netflix, Yahoo, Tripadvisor, Last.fm ve IMDb gibi popüler internet sitelerinde önemli bir yer tutar. Aynı zamanda birçok medya şirketi de abonelerine sundukları hizmetlerde öneri sistemi kullanır. Örneğin, Netflix, öneri sisteminin performansını geliştiren ilk takıma bir milyon dolarlık ödül vermiştir. (Lohr, 2009)

Öneri sistemlerinin başarısını değerlendirmek için doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metrikler kullanılmaktadır. Bunun yanı sıra, kullanıcı memnuniyeti anketleri ve geri bildirimler de performans değerlendirmesinde önemli rol oynamaktadır.

Öneri sistemleri içerik tabanlı, işbirlikçi filtreleme tabanlı, demografik bilgi tabanlı, bilgi tabanlı, topluluk tabanlı ve karma öneri sistemleri olarak kategorize edilmektedir (Ricci vd., 2010). İçerik tabanlı öneri sistemleri, kullanıcının geçmişte beğendiği öğelere benzeyen yeni öğeleri önermeyi öğrenir. Bu benzerlik, öğelerin özelliklerine dayalı olarak hesaplanır. Örneğin, bir kullanıcı komedi filmlerini olumlu derecelendirdiyse, sistem benzer türdeki diğer filmleri önerebilmektedir (Ricci vd., 2010).

İşbirlikçi filtreleme tabanlı öneri sistemleri, kullanıcılara, benzer zevklere sahip diğer kullanıcıların beğendiği öğeleri önerir. Kullanıcılar arasındaki benzerlik, derecelendirme geçmişlerine dayalı olarak hesaplanmıştır. Bu yaklaşım, "insanlar arasındaki ilişki" olarak da adlandırılır ve öneri sistemlerinde yaygın olarak kullanılan tekniktir (Ricci vd., 2010).

Demografik bilgi tabanlı öneri sistemleri, kullanıcının demografik özelliklerine göre öneriler sunar. Farklı demografik grupların farklı önerilere ihtiyaç duyduğuna dayanarak çalışır. Web siteleri, dil veya ülke bilgilerine göre kişiselleştirilmiş öneriler sunabilir. Bu yaklaşımlar pazarlamada yaygın olsa da demografik sistemler üzerine yapılan araştırmalar sınırlıdır (Ricci vd., 2010).

Bilgi tabanlı öneri sistemleri, belirli bir konuya dair sahip oldukları bilgiye dayanarak öğe önerir. Özellikle durum temelli sistemler, kullanıcı ihtiyaçları ile öneriler arasındaki benzerliği tahmin eden benzerlik fonksiyonları kullanır. Bu benzerlik skoru, önerinin kullanıcıya olan faydasını göstermektedir (Ricci vd., 2010).

Topluluk tabanlı öneri sistemleri, kullanıcıların arkadaşlarının tercihlerini temel alarak öneriler sunar. Araştırmalar, insanların arkadaşlarının önerilerine daha fazla güvendiğini göstermektedir. Bu sistemler, sosyal ağlar aracılığıyla kullanıcıların sosyal ilişkilerini ve arkadaşlarının tercihlerine dair verileri toplar ve önerilerini bu bilgilere dayandırmaktadır (Ricci vd., 2010).

Karma öneri sistemleri, birden fazla öneri tekniğini birleştirir. Farklı tekniklerin avantajlarını kullanarak dezavantajları gidermeye çalışır. Örneğin, işbirlikçi filtreleme yeni öğelerle ilgili sorun yaşayabilirken, içerik tabanlı yaklaşımlar öğelerin özelliklerine dayalı tahminler yapabilir. İki veya daha fazla öneri sistemi tekniğinin birleştirilmesiyle karma bir sistem oluşturulabilmektedir (Ricci vd., 2010).

3.2. Büyük Dil Modeli

Dil modelleri (Language Model, LM), insan dilini anlama ve üretme yeteneğine sahip hesaplamalı modellerdir. En yaygın tür olan n-gram modelleri, önceki bağlama dayalı olarak kelime olasılıklarını tahmin etmektedir (Chang vd., 2023).

Son yıllarda, Üretken önceden eğitilmiş dönüştürücüler (Generative Pre-trained Transformer, GPT) gibi büyük dil modelleri (Large Language Model, LLM), doğal dil işleme (Natural Language Processing, NLP) alanında büyük ilerlemeler sağlamıştır. Bu modeller, geniş metin verileriyle eğitilerek, insan benzeri metinler üretme, soruları yanıtlama ve dil görevlerini yüksek doğrulukla gerçekleştirme yeteneğine sahiptir (Kasneci vd., 2023). LLM'ler, bağlam veya önerme temel alarak metin üretme yeteneği sayesinde daha tutarlı ve uygun yanıtlar verebilir, bu da onları interaktif uygulamalar için ideal kılmaktadır. İnsan geri bildiriminden takviyeli öğrenme (Reinforcement Learning from Human Feedback, RLHF) tekniği, LLM'lerin hatalarından öğrenerek performansını geliştirmesine yardımcı olmaktadır (Chang vd., 2023).

LLM'ler genel olarak açık kaynaklı (open source) ve kapalı kaynaklı (closed source) modeller olarak ikiye ayrılır. Açık kaynaklı modeller, kontrol noktalarının (checkpoints) genel erişime açık olduğu modellerdir. Araştırmacılar, performansı artırmak için bu modelleri öneri verileriyle ince ayar yaparlar. LLaMA ve Vicuna gibi açık kaynaklı modeller, öneri sistemlerinde yaygın olarak kullanılır. Parametre-etkin ince ayarlama (Parameter-efficient fine-tuning) stratejilerinde, düşük sıra uyarlama (Low-Rank Adaptation, LoRA) gibi yöntemler tercih edilir. Google'ın Flan-T5 ve Tsinghua Üniversitesi'nin ChatGLM modelleri de bu alanda popülerdir. Açık kaynaklı modellerin erişilebilir kontrol noktaları, öneri görevlerine özgü parametre ayarlamalarına imkân tanımaktadır (Lanling vd., 2023).

Kapalı kaynaklı modeller ise kontrol noktalarına genel erişimin olmadığı modellerdir. Bu tür modellerde, soğuk başlangıç (cold start) senaryolarında sıfır atışlı (zero-shot) öneri yetenekleri genellikle incelenir. En yaygın kapalı kaynaklı model, OpenAI'in ChatGPT serisidir. Daha güçlü olan GPT-4, hala açık kaynaklı değildir. Kapalı kaynaklı modellerde

kontrol noktalarına erişim olmadığı için, OpenAI, model performansını iyileştirmek amacıyla web eklentileri ve ayarlama arayüzleri gibi araçlar sunar. Ancak, bu modellerin esnekliği, yüksek maliyet ve kapalı kutu (Black Box) özellikleri nedeniyle sınırlıdır. Bu zorlukları aşmak için, mevcut literatürde öneri tasarımı (recommendation design), geri alım geliştirme (feedback enhancement) ve geleneksel öneri modellerinin kombinasyonu gibi yöntemler kullanılarak kapalı kaynaklı modellere öneri bilgisi eklenmektedir (Lanling vd., 2023).

3.3. Komut Mühendisliği

Bir komut (prompt), bir LLM'e verilen talimatların bir kümesidir ve LLM'i özelleştirmek, geliştirmek veya yeteneklerini iyileştirmek için kullanılır. Bir komut, belirli kurallar ve yönergeler sağlayarak LLM ile sonraki etkileşimleri ve çıktı üretimini etkilemektedir (White vd., 2023)

Temelde, bir komut, konumu belirler, önemli bilgileri tanımlar ve istenen çıktı biçimini ve içeriğini belirler. Örneğin, bir komut, LLM'in belirli bir kodlama stiline uymasını veya oluşturulan bir belgede belirli anahtar kelimeleri işaretlemesini sağlayabilir. Bu yönergeler, komutların daha yapılandırılmış ve incelikli çıktıları kolaylaştırmasını sağlar ve LLM'in bağlamında çeşitli yazılım mühendisliği görevlerine yardımcı olmaktadır.

Komut mühendisliği (prompt engineering), kullanıcıların LLM'i yönlendirmek için özel öneri metinleri oluşturduğu yaygın bir yaklaşımdır. LLM'lerin dönüştürücü (transformer) mimarisi, bağlamsal öğrenme ve RLHF yetenekleri, doğal dil işleme alanında önemli bir ilerleme kaydetmiştir.

Komut mühendisliği, LLM'in komutlar aracılığıyla programlanma sürecidir ve bu süreç beş kategoriye ayrılır: giriş anlamı, çıktı özelleştirme, hata tanımlama, komut iyileştirme ve etkileşim.

Giriş anlamı kategorisi, LLM'in girdiyi nasıl anladığı ve bunu çıktıya dönüştürdüğüyle ilgilenir. Bu kategoride, LLM'in anlaması için özel bir dil oluşturmayı amaçlayan meta dil oluşturma kalıbı bulunmaktadır (White vd., 2023).

Çıktı özelleştirme kategorisi, LLM tarafından üretilen çıktının türünü, formatını veya yapısını düzenler. Bu kategoride çıktı oluşturucu, karakter, görselleştirme oluşturucu, tarif ve şablon gibi kalıplar yer alır. Bu kalıplar, otomatik komut dosyası oluşturma, karakter belirleme, görselleştirme yapma, belirli bir sonuca ulaşma adımları oluşturma ve şablonlarla çalışmayı içermektedir (White vd., 2023).

Hata tanımlama kategorisi, LLM çıktısındaki hataları belirleyip düzeltmeyi amaçlar. Gerçek kontrol listesi, çıktıdaki bilgilerin doğruluğunu kontrol ederken, yansıma kalıbı LLM'in kendi çıktısını değerlendirmesini sağlamaktadır (White vd., 2023).

Komut iyileştirme kategorisi, girdi ve çıktı kalitesini artırmaya odaklanır. Soru düzeltme, alternatif yaklaşımlar, bilişsel doğrulayıcı ve reddi kıran gibi kalıplar; soruların daha iyi versiyonlarını önermek, alternatif çözüm yolları sunmak, alt sorular üretmek ve reddedilen soruları yeniden ifade etmek gibi işlevler görmektedir (White vd., 2023).

Etkileşim kategorisi, kullanıcı ve LLM arasındaki etkileşim dinamiklerini düzenler. Bu kategori, LLM'in sorular sormasını, oyun formatında çıktı üretmesini veya sürekli olarak çıktı üretmesini sağlayan ters etkileşim, oyun oynama ve sonsuz üretim kalıplarını içermektedir (White vd., 2023).

Son olarak, bağlam kontrolü kategorisi, LLM'in işlediği bağlamsal bilgileri yönetir ve bağlam yöneticisi kalıbı ile kullanıcının LLM'in çıktısı için bağlamı belirlemesine olanak tanımaktadır (White vd., 2023).

Komut mühendisliği kapsamında geliştirilen bir fonksiyon, komut tipine göre (ör. zero_shot, few_shot, role_play, chain_of_thought) bağlı olarak sisteme ek bağlam eklemektedir.

- Sıfır atışlı (Zero shot, temel): Sistemin yalnızca verilen komutu anlayarak bir bağlam olmadan yanıt üretmesi.
- Sıfır atışlı rol yapma (Zero shot role play, başarılı bir şef gibi rol yapma): Sistemden, belirli bir uzman (ör. başarılı bir şef) gibi davranarak yanıt vermesinin istendiği durum.
- Sıfır atışlı düşünce zinciri (Zero shot chain of thought, adım adım düşünüp gerekçeleri açıklama): Sistemden, verdiği yanıtları adım adım gerekçelendirerek açıklaması beklenen durumu ifade etmektedir.
- Az atışlı (Few shot, kullanıcının önceki yüksek puanlı tariflerinden yararlanma): Sistem, önceki yüksek puanlı tarif örneklerini kullanarak öneriler oluşturmaktadır.
- Az atışlı rol yapma (Few shot role play, hem önceki tariflerden yararlanma hem de rol yapma): Sistemden, hem kullanıcıya ait önceki yüksek puanlı tariflerden yararlanarak hem de bir uzman gibi davranarak yanıt üretmesi istenen durumu ifade etmektedir.
- Az atışlı düşünce zinciri (Few shot chain of thought, hem önceki tariflerden yararlanma hem de gerekçeli adım adım düşünme): Sistemden, önceki tariflerden yararlanarak adım adım gerekçeli bir şekilde yanıt üretmesi beklenen durumu ifade etmektedir.

Fonksiyon, temel bir çerçeve (base_context) ve seçilen komut tipine özgü ek betimleyici cümleler (context_additions) birleştirerek, modelin hangi rolü üstleneceğini ve nasıl davranması gerektiğini ilgili komuta yansıtmaktadır.

Komut algoritmasının son bolümü olan bu kısımda, sistemden hangi adımları izlemesinin beklendiği, hangi kriterlere göre puanlama yapacağı ve nasıl bir çıktı formatı sunacağı anlatılır, parametrik olarak belirlenen komut parçaları bir araya getirilerek bütün bir komut metni oluşturulmaktadır.

3.4. Öneri Sistemlerindeki Komut Stratejileri

Komut stratejisi, büyük dil modellerinin (LLM'ler) belirli bir görevi yerine getirmesi için verilen talimatlar veya ipuçlarını içeren bir yöntemdir. Bu strateji, modelin nasıl yanıt vermesi gerektiğini yönlendirir ve istenen sonuçları elde etmek için uygun bir bağlam, örnekler veya açıklamalar sunar. Komut stratejileri, modelin daha doğru ve alakalı yanıtlar üretmesine yardımcı olmak amacıyla tasarlanmıştır.

Yönlendirme stratejileri, LLM'lere referans örnekleri olmaksızın doğrudan bağlam bilgisi ve görev açıklamaları sunarak temel bir yaklaşım oluşturur. Bu strateji, görev açıklamaları ve yönlendirme formatlarının öneri görevlerine göre değişiklik gösterebilmesini sağlar (Lanling vd., 2023).

Az atışlı komut (few-shot prompt) stratejisi, LLM'in kullanıcı niyetlerini daha iyi anlaması için birkaç örnek içerir. Bu yaklaşım, sağlanan örneklerin kişiselleştirilmiş kullanıcı ilgi alanlarıyla uyumlu olması durumunda etkilidir (Lanling vd., 2023).

Yakın zaman odaklı komut (recency-focused prompt), son etkileşimde bulunulan öğelerin daha güçlü bir ilişkiye sahip olduğu gözlemiyle, bu öğelere vurgu yapar (Lanling vd., 2023).

Rol yapma tabanlı komut (role play prompt) stratejisi, dil modeliyle rol yapma oyunu formatında etkileşimi içerir ve öneri sistemlerinin belirli hedeflerle kullanıcıya hizmet etmesini sağlar (Lanling vd., 2023).

Düşünce zinciri komut (chain-of-thought prompt), mantık soruları gibi görevlerde, LLM'in adım adım sorunları çözmesine veya görevin mantığını ve analiz sürecini açığa çıkarmasına yardımcı olur (Lanling vd., 2023).

Kendi kendine komut (self-prompting) stratejisi, LLM'in soruları yanıtlamak için kendi kendine bilgi üretmesini ve bu bilgiyi kullanarak sonuçlar elde etmesini sağlar.

Genellikle, farklı çıkarım zincirlerinin sonuçları üzerinde çoğunluk oylaması yapılarak nihai tahminler belirlenir (Lanling vd., 2023).

Özetle, bu farklı yönlendirme stratejileri, öneri sistemlerinde farklı amaçlara hizmet eder ve doğru kullanıldığında sistemlerin performansını artırabilir (Lanling vd., 2023).

3.5. Yemek Tarifi Öneri Sistemleri

Yemek öneri sistemleri, kullanıcıların tercihlerine uygun tarifler sunarak onların memnuniyetini artırmayı hedefleyen dijital araçlardır. Bu sistemler, tarifleri kaloriler ve mutfak türlerine göre sınıflandırarak hem bireysel tercihlere hem de sağlık odaklı seçimlere uygun öneriler sunabilmektedir. İşbirlikçi filtreleme (Collaborative Filtering, CF) ve içerik tabanlı filtreleme (Content-Based Filtering, CB) gibi yöntemlerin yanı sıra, bu yaklaşımları birleştiren hibrit sistemler de kullanılmaktadır (Jhamat vd., 2021).

CF, kullanıcıların geçmişte beğendiği tarifleri analiz ederek benzer tercihlere sahip diğer kullanıcıların beğenilerini öngörürken, CB tariflerin özelliklerine odaklanarak kişiselleştirilmiş öneriler sunmaktadır. Bununla birlikte, CF yöntemi "seyreklik" (sparsity) ve "soğuk başlangıç" (cold start) gibi sorunlarla karşı karşıya kalabilirken, CB yöntemi kullanıcının profilini fazla daraltarak önerilerin çeşitliliğini sınırlayabilir. Bu zorlukları aşmak için geliştirilen Hibrit Hiyerarşik Küme Öneri Sistemi (Hybrid Hierarchical Clusters Recommender System, HHCRS), tarifleri öncelikle mutfak türüne, ardından kalori değerine göre gruplandırarak daha hassas önerilerde bulunmayı mümkün kılmaktadır (Jhamat vd, 2021).

HHCRS'nin, tariflerin kaloriler ve mutfak türlerine göre sınıflandırılmasının öngörü doğruluğunu artırdığı ve seyreklik sorunlarını azalttığı gösterilmiştir. Ancak, bu yaklaşımın seyrek veri setleri üzerinde daha karmaşık modellerle desteklenmesi gerektiği vurgulanmaktadır. Gelecekteki çalışmalar, kullanıcı mahremiyetini koruyarak daha geniş veri setleriyle bu sistemlerin performansını iyileştirmeyi hedefleyebilir (Jhamat vd., 2021).

3.6. Komut Oluşturma Algoritması

Parametrik komut üretimi, bu çalışma kapsamında geliştirilmiş olan kullanıcıya veya sisteme özel gereksinimler doğrultusunda otomatikleştirilmiş ve uyarlanabilir bir şekilde komut (prompt) oluşturulmasını sağlayan bir algoritmadır. Günümüzde, birçok içerik üretiminde (metin tabanlı yaratıcı yazarlık, soru yanıtlama, öneri sistemleri vb.) yapay zekâ modellerinin, özellikle büyük dil modellerinin (LLM), kullanımı yaygınlaşmıştır (Xu vd., 2018). Bu modellerin doğru ve tutarlı çıktı üretebilmesi, girilen komutların kalitesine ve içeriğine bağlıdır.

Geliştirilen algoritma temel olarak kullanıcının belirttiği bir dizi anahtar malzemeyi temel alarak, mevcut bir tarif listesi içinden en uygun tarifleri bulmayı ve kullanıcının daha önce sevdiği (top-rated) tariflerden de yararlanarak yüksek puan alabilecek yeni tarifler önermeyi hedefler.

Bu amaçla, komut tipi (prompt_type) gibi parametrik seçimler mevcuttur. Her komut tipi, LLM'nin nasıl yanıt vereceğini, ne kadar ayrıntılı düşüneceğini (chain of thought), rol yapıp yapmayacağını (role_play) ve kullanıcının geçmişte en beğendiği tariflerden yararlanıp yararlanmayacağını (few_shot) belirler. Parametrik yaklaşım, kodu tekrar kullanılabilir ve genişletilebilir hâle getirir. Farklı parametreler değiştirilerek, farklı senaryolara uygun istemler üretmek mümkündür.

Kullanılan parametreler ve LLM'e komut içerisinde gönderilen komut biçimi yönlendirmeleri aşağıdaki gibidir:

- LLM'in önermesi beklenen tarif adedi (num_recipes) kadar öneri talep edilmektedir.
- En az bir anahtar malzeme (örneğin tavuk, patates gibi) içermesi beklenmektedir.
- Yeni önerilecek tarifleri geçmiş kullanıcı tercihlerine göre 0–5 arasında puanlayıp en yüksek puanlılar üst sırada olacak şekilde sıralaması istenmektedir.
- Metin çıktı formatı detaylı şekilde belirtilmektedir.

- Komut tipi eğer “zero_shot” içeriyorsa, “önceki en yüksek puanlı tarifler” bölümü komut içerisinden kaldırılmaktadır (çünkü zero-shot senaryoda geçmiş veriye referans yapılmamaktadır).

3.7. Önerilen Yaklaşım

Yemek tarifi öneri sistemi, kullanıcıların kişiselleştirilmiş tarif önerileri alabilmesi için yapılandırılmış bir süreç sunmaktadır. Sistem, kullanıcının kendine ait kullanıcı adı ve şifresiyle giriş yapmasıyla başlar. Kullanıcı, sistemde ilk adım olarak tarif almak istediği yemek türünü seçer. Bu adım, tarifin ana kategorisinin belirlenmesini sağlar ve sistemin kullanıcıya uygun içerikler sunabilmesi için temel bir başlangıç noktası oluşturur.

Ardından kullanıcı, tarifte kullanılmasını istediği malzemeleri seçer. Bu seçimler hem kullanıcı tercihlerine hem de tarif önerilerinin çeşitliliğine yön verir.

Sistem, kullanıcıların daha önce yemek tarifi değerlendirmesi yapıp yapmadığını dikkate alır. Eğer kullanıcı daha önce tarif değerlendirmişse, sistem kullanıcı geçmişini analiz ederek seçilen malzemeler, yemek tipi ve önceki tercihleri doğrultusunda en beğenilen N adet yemeği belirler. Bu süreç, kullanıcının geçmiş deneyimlerinden yola çıkarak daha kişisel ve ilgili öneriler sunmayı amaçlar.

Ardından, sistem bu verilere dayalı olarak kullanıcının tercihlerine uygun bir tarif oluşturmak için bir komut hazırlar. Bu komut, kullanıcı verilerini ve önceki beğenileri dikkate alarak LLM’e gönderilir ve kişiselleştirilmiş bir tarif önerisi oluşturulur. Uygulamaya ait arayüzün ekran görüntüsü Şekil 3.1’de verilmiştir.

Diğer yandan, kullanıcı daha önce herhangi bir yemek tarifi değerlendirmesi yapmamışsa, sistem genel kullanıcı verilerini dikkate alır. Bu durumda, seçilen malzemeler, eşleşmeleri ve yemek tipine göre diğer kullanıcılar tarafından en beğenilen N adet yemek belirlenir. Böylece, kullanıcı tercihi bulunmayan durumlarda bile sistem, popüler tariflerden yola çıkarak öneriler sunar.

Bu bilgiler ile tüm kullanıcı tercihleri parametre olarak kullanılarak bir komut hazırlanmakta ve LLM'e iletilmektedir.

Son aşamada, sistem tarafından hazırlanan komut LLM'e gönderilir ve kullanıcının seçtiği malzemeler ve yemek tipi doğrultusunda LLM'den bir yemek tarifi önerisi alınır. LLM'den alınan yanıt, kullanıcının beğenisine sunulur ve süreç tamamlanır. Bu yapılandırılmış süreç hem kullanıcı geçmişini hem de tüm kullanıcı tercihlerini dikkate alarak kişiselleştirilmiş ve çeşitli tarif önerileri sunmayı hedeflemektedir.



Şekil 3.1. Uygulama arayüzü

Önerilen yaklaşıma ait akış şeması Şekil 3.2'de verilmiştir.

Şekil 3.2'de verilen akış şemasında yeşil ile renklendirilmiş dikdörtgen şeklindeki adımlar kullanıcı tarafından bir seçim yapılarak ya da bir butona tıklanarak gerçekleştirilen interaktif işlemleri; arka plan rengi koyu gri olan paralel kenar şeklindeki adımlar, bir önceki adımın sonucu olarak ortaya çıkan verileri; arka planı açık gri ile renklendirilmiş adımlar kullanıcı tarafından bir işlem beklenmeden arka planda çalışan algoritmanın otomatik gerçekleştirdiği işlemleri ve son olarak arka planı siyah olarak renklendirilmiş eşkenar dörtgen ise bir karar noktasını ifade etmektedir.

Akış kullanıcının daha önce sisteme giriş yapıp yapmadığı detayı özelinde ikiye ayrılmakta ve bu iki kısım farklı renklendirilmiş alanlar ile ifade edilmiştir.

Sol tarafta kalan açık mavi ile renklendirilmiş alan daha önce sisteme giriş yapmamış ya da herhangi bir değerlendirmede bulunmamış kullanıcıları ifade ederken sağ tarafta kalan sarı ile renklendirilmiş kısım ise sistemde en az bir yemek için değerlendirme puanı bulunan kullanıcılar için akışı ifade etmektedir.

Bu akışlar son adımda birleşmekte ve sonuç olarak oluşturulan komut LLM'e gönderilmektedir.

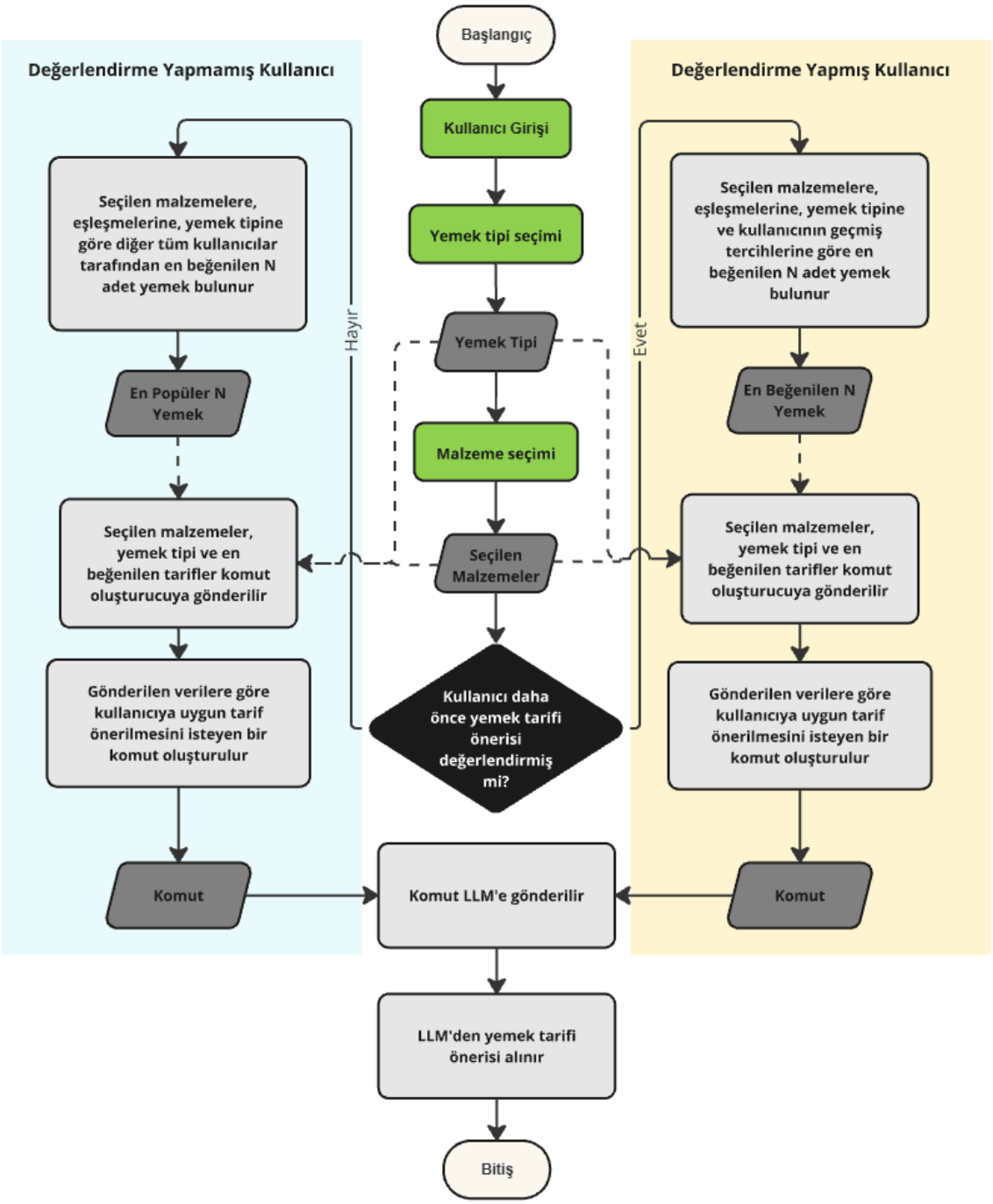
Şekil 3.2'de görülen akış şemasındaki kesikli çizgiler bir adımdan sonrakine bir veri transferi gerçekleştiğini, düz çizgiler ise direkt olarak sonraki adıma geçildiğini işaret etmektedir.

3.8. Önerilen Yaklaşımın Performans Değerlendirmesi

Önerilen yaklaşımın performansının değerlendirilmesi için aşağıdaki adımlar uygulanmıştır:

Adım 1: Kullanıcının giriş yapmasıyla algoritma işlemeye başlar. Giriş işlemi, kişiselleştirilmiş tarif önerilerinin oluşturulması için gereklidir ve sistemin kullanıcıyı

tanıyarak ilgili veri tabanlarını ve algoritmaları çalıştırmasını sağlar. Kullanıcı giriş yaptıktan sonra, süreç tamamen sistem tarafından otomatik olarak yürütülmektedir.



Şekil 3.2. Önerilen yaklaşım akış şeması

Adım 2: Kullanıcının giriş yapmasının ardından, sistem malzeme seçimini otomatik olarak gerçekleştirir. Bu işlem, yemek tarifleri veri tabanına dayanır ve iki yöntemle

yapılır. Sistem en sık kullanılan 50 malzemeden rastgele N adet malzeme seçer ya da veri tabanındaki en sık kullanılan N malzemeyi doğrudan belirler. Bu seçim, tarif önerilerinin çeşitli ve kullanıcı ihtiyaçlarına uygun olmasını sağlamayı amaçlamaktadır.

Adım 3: Malzeme seçimi tamamlandıktan sonra sistem, tarif oluşturma algoritmasında kullanılacak parametrelere göre yapılacak değişiklikleri komut üzerine uygular. Bu parametreler, önceden tanımlı senaryo setlerine dayalıdır ve LLM'in nasıl çalışacağını belirler. Her bir senaryo seti, model sürümü, sıcaklık, komut tipi ve önerilecek tariflerin sayısı gibi ayarları içermektedir. Sistem, bu parametreleri bir döngü yapısı içinde işleyerek her senaryo için ayrı bir işlem başlatmaktadır.

Adım 4: Her bir senaryo için sistem, seçilen malzemeler ve parametre setine dayalı olarak bir komut oluşturur. Bu komut, LLM'in tarif önerilerini belirlenen kriterlere uygun şekilde üretmesi için gerekli talimatları içerir. Döngü yapısı sayesinde, sistem tüm senaryolar için ayrı ayrı komutlar oluşturarak süreç tamamlanmaktadır.

Adım 5: Ouşturulan komutlar bir bütün olarak LLM'e iletilmektedir. Sistem, her bir senaryo için LLM'in doğal dil işleme yeteneklerinden faydalanarak tarif önerilerini alır. LLM'e gönderilen komutlar, tariflerin kullanıcı ihtiyaçlarına ve belirlenen parametrelere uygun şekilde hazırlanmasını sağlamaktadır.

Adım 6: LLM, gönderilen her bir komut için tarif önerileri üretmektedir. Bu aşamada sistem herhangi bir ek işlem yapmaz; yalnızca LLM'den alınan sonuçları kaydeder. Her senaryo için ayrı ayrı alınan tarif önerileri, sistemde saklanarak gelecekteki süreçler için kullanılmak üzere veri tabanına kaydedilmektedir.

Örnek bir komut ve LLM tarafından verilen yanıt Şekil 3.3 ve Şekil 3.4'te verilmiştir. Şekil 3.3 içerisindeki ilgili alanlarda, gönderilen komutun parametrik olarak değişebilen alanları aşağıdaki şekilde ile gösterilmiştir:

- Arasından yemek tarifi seçilecek ve sıralanacak yemek tariflerinin listesi mavi renk ile gösterilmiştir.

- Seçilecek yemek tariflerinde en az birinin bulunması istenen temel malzemeler yeşil renk ile gösterilmiştir.
- LLM tarafından önerilmesi beklenen tarif adedi kırmızı renk ile gösterilmiştir.

1.Komut Parçası: Görev Bağlamı
You are a content-based recipe recommender system with expert-level cooking knowledge. *Your task is to recommend recipes that align with the user's preferences, based on user-provided ingredients.* *Think step-by-step and consider how each ingredient or recipe element influences the user preference score.* *Explain your reasoning in detail as you proceed.* *You are given a list of available recipes: (onion herb bread bread machine, spiced basmati rice, triberry muffins, chocolate anise cookies biscotti, broussard s shrimp chandeleur, triple chocolate espresso bean cookies, bread pudding with raspberry strawberry topping, cranberry lemon cupcakes, rice with carrots and peas rice cooker, poached cod with lemon butter sauce, coffee braised pot roast with onion sauce, chicken breasts with broccoli cheese, moist tender spice cake gingerbread variation, orange tea biscuits, black bean pumpkin soup, espresso chocolate chip muffins, saumons smoked en pte feuillete, chocolate chip scones, apple and cheese scones, finest vanilla cupcakes).* *You must use these recipes, do not change their name, do not use or add any other recipes which is not given in this list.- A list of key ingredients to prioritize: # salt, sugar, butter, onion, water #.*
2.Komut Parçası: ChatGPT'ye Verilen Görev
1. Select and recommend exactly <num_recipes> recipe names as a random recipe type, from the given list that you believe will be top-rated by this user. *- Recipes must align with user preferences and contain at least one key ingredient.* *- If there are no recipes likely to receive a perfect score of 5, choose those that are closest to getting 5 out of 5 rating.* *- Prioritize recipes with multiple key ingredients, ranking those highest.* *- Exclude recipes without any key ingredients.* *2. Assign ratings (0-5) to each recipe based on how well it matches user preferences.* *- 5: The user is very likely to love it.* *- 0: The user would never like it.* *- Choose and recommend only the recipes which are most-likely to be top-rated by this user.* *- Criteria:* *1. Ingredient match: More key ingredients mean higher ratings.* *2. Novelty: Do not copy previously top-rated recipes directly; recommendations must come from the provided list.* *3. Sort recommendations in descending order by rating. For ties, prioritize recipes with more key ingredients.* *Provide recommendations in this format:* *| Recipe Name | Rating |* *|----------|--------|* *| #<recipe name># | %<rating>% |*
3.Komut Parçası: Kısıtlar
Constraints: *- Only use recipes from the provided list; do not recommend new ones.* *- Enclose recipe names in # and ratings in %.* *- Never alter or omit text features such as capitalization in the recipe list provided to you; use them in your recommendations exactly as they are given in the available recipes list.* *- Never correct errors in given recipe names, such as words, excessive spaces, punctuation, etc. Use them in your recommendations exactly as they are given in the available recipes list.*

Şekil 3.3. Verilen algoritma ile oluşturulan bir örnek az atışlı rol yapma komutu

Gönderilen Tam Komut
You are a content-based recipe recommender system with expert-level cooking knowledge. *Your task is to recommend recipes that align with the user's preferences, based on user-provided ingredients.* *Think step-by-step and consider how each ingredient or recipe element influences the user preference score.* *Explain your reasoning in detail as you proceed.* *You are given a list of available recipes: (onion herb bread bread machine, spiced basmati rice, triberry muffins, chocolate anise cookies biscotti, broussard s shrimp chandeleur, triple chocolate espresso bean cookies, bread pudding with raspberry strawberry topping, cranberry lemon cupcakes, rice with carrots and peas rice cooker, poached cod with lemon butter sauce, coffee braised pot roast with onion sauce, chicken breasts with broccoli cheese, moist tender spice cake gingerbread variation, orange tea biscuits, black bean pumpkin soup, espresso chocolate chip muffins, saumons smoked en pte feuillete, chocolate chip scones, apple and cheese scones, finest vanilla cupcakes).* *. You must use these recipes, do not change their name, do not use or add any other recipeswhich is not given in this list.- A list of key ingredients to prioritize: # salt, sugar, butter, onion, water #.* *1. Select and recommend exactly 10 recipe names as a random recipe type, from the given list that you believe will be top-rated by this user.* *- Recipes must align with user preferences and contain at least one key ingredient.* *- If there are no recipes likely to receive a perfect score of 5, choose those that are closest to getting 5 out of 5 rating.* *- Prioritize recipes with multiple key ingredients, ranking those highest.* *- Exclude recipes without any key ingredients.* *2. Assign ratings (0-5) to each recipe based on how well it matches user preferences.* *- 5: The user is very likely to love it.* *- 0: The user would never like it.* *- Choose and Recommend only the recipes which are most-likely to be top-rated by this user.* *- Criteria:* *1. Ingredient match: More key ingredients mean higher ratings.* *2. Novelty: Do not copy previously top-rated recipes directly; recommendations must come from the provided list.* *3. Sort recommendations in descending order by rating. For ties, prioritize recipes with more key ingredients.* *Provide recommendations in this format:* *| Recipe Name | Rating |* *|----------|--------|* *| #<recipe name># | %<rating>% |* *Constraints:* *- Only use recipes from the provided list; do not recommend new ones.* *- Enclose recipe names in # and ratings in %.* *- Never alter or omit text features such as capitalization in the recipe list provided to you; use them in your recommendations exactly as they are given in the available recipes list.* *- Never correct errors in given recipe names, such as words, excessive spaces, punctuation, etc. Use them in your recommendations exactly as they are given in the available recipes list.*

Şekil 3.3. Verilen algoritma ile oluşturulan bir örnek az atışlı rol yapma komutu (devam)

Bütün parametreler bir araya getirilerek bir komut halinde LLM'e gönderildikten sonra Şekil 3.4'teki yanıt alınmaktadır.

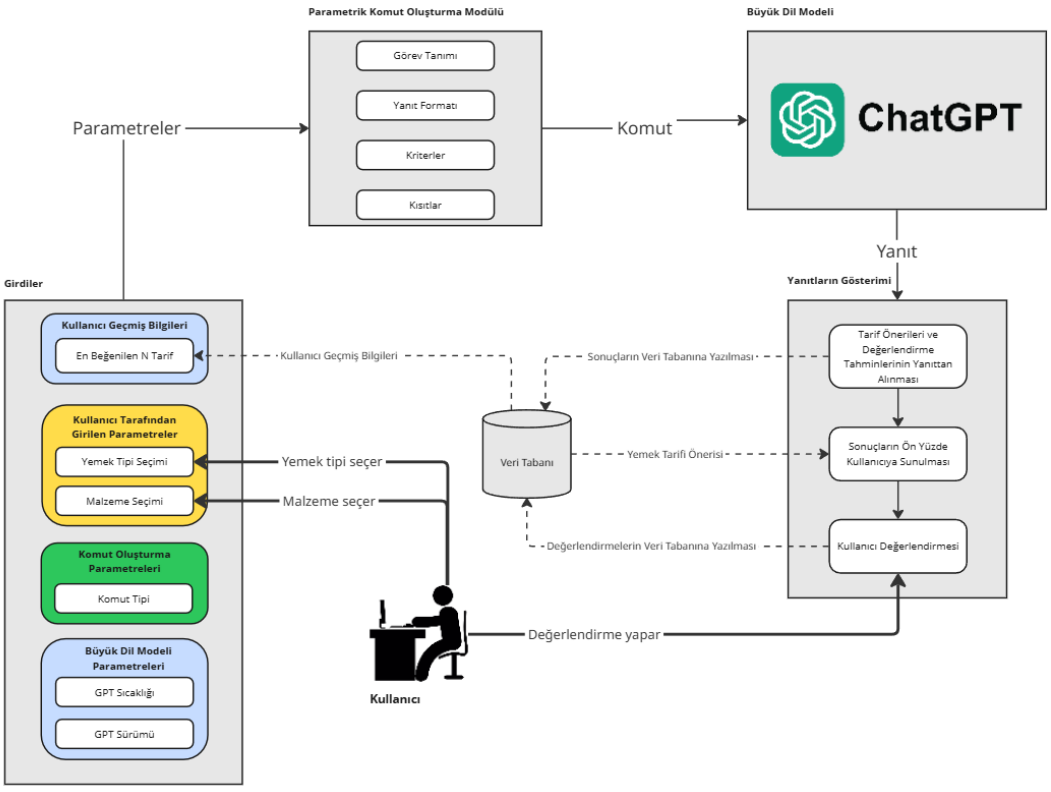
"To recommend the best recipes based on the provided ingredients, let's evaluate each recipe for the presence of key ingredients (salt, sugar, butter, onion, water). We will prioritize recipes that contain multiple key ingredients and assign ratings accordingly, ensuring a mix of novelty and strong ingredient alignment.

Recommendations

| Recipe Name | Rating |
|----------|--------|
| #coffee braised pot roast with onion sauce# | %5% |
| #poached cod with lemon butter sauce# | %5% |
| #onion herb bread bread machine# | %4% |
| #spiced basmati rice# | %4% |
| #black bean pumpkin soup# | %4% |
| #chicken breasts with broccoli cheese# | %4% |
| #rice with carrots and peas rice cooker#| %4% |
| #cranberry lemon cupcakes# | %3% |
| #moist tender spice cake gingerbread variation# | %3% |
| #chocolate chip scones# | %3% |"

Şekil 3.4. Verilen algoritma ile gönderilen komut için alınan örnek bir yanıt

Önerilen yaklaşıma ait sistem modülleri, kullanıcı etkileşimleri ve LLM arasındaki akış Şekil 3.5'te gösterildiği gibidir.



Şekil 3.5. Önerilen yaklaşıma ait sistem modülleri, kullanıcı etkileşimleri ve LLM arasındaki akış

3.9. Performans Değerlendirme Ölçütleri

Önerilen yaklaşımı değerlendirmek üzere kullanılmış ölçütler aşağıdaki gibidir:

Normalize edilmiş indirgenmiş kümülatif kazanç (Normalized Discounted Cumulative Gain, NDCG), önerilen tariflerin sıralama doğruluğunu ölçmek için kullanılır. Kullanıcıların geçmiş değerlendirmeleri ile öneri sıralamasının uyumu analiz edilir.

Denklem 3.1'de verilen DCG; LLM'lerin sıralama ve sıralanan ögelerin alaka düzeylerine göre sıralanıp sıralanmadığını ölçen, literatürde yaygın olarak kullanılan bir ölçüttür. (Xu vd., 2018)

$$DCG = \frac{(2^{\hat{y}} - 1)}{\log_2(\hat{r} + 1)} \tag{3.1}$$

Denklem 3.1'de verilen DCG değişkeni indirgenmiş kümülatif kazanç (Discounted Cumulative Gain) değerini, $\hat{y}$ değeri LLM tarafından tahmin edilen değerlendirme puanını, $\hat{r}$ değişkeni ise LLM tarafından tahmin edilen sıralama değerini ifade etmektedir.

$$iDCG = \frac{(2^{y_{ideal}} - 1)}{\log_2(r_{ideal} + 1)} \tag{3.2}$$

Denklem 3.2'de görülen $iDCG$ değişkeni ideal indirgenmiş kümülatif kazanç (Ideal Discounted Cumulative Gain) değerini, y_{ideal} değişkeni ideal değerlendirme değerini ve r_{ideal} parametresi ideal sıralama değerini ifade etmektedir.

$$NDCG = \frac{DCG}{iDCG} \tag{3.3}$$

Denklem 3.3'te verilen $NDCG$ değişkeni normalize indirgenmiş kümülatif kazanç (Normalized Discounted Cumulative Gain) değerini ifade etmektedir.

WNDCG, yani Ağırlıklandırılmış Normalize İndirgenmiş Kümülatif Kazanç, NDCG'nin ağırlıklı versiyonudur. Bu metrik, kullanıcının geçmişte beğendiği tarifler veya tercihleri

gibi farklı faktörlerin ağırlıklarını göz önünde bulundurur (Tamm vd., 2021). Farklı çalışmalarda çeşitli ağırlıklandırma yöntemleri kullanılsa da bu çalışmada önerilen WNDCG formülüne literatürde rastlanmamıştır.

Mevcut ölçütler, kullanıcı değerlendirme puanlarını ve önerilerin sıralamasını aynı anda değerlendirmekte yetersiz kaldığından, bu yaklaşım yeni ve daha kapsamlı bir performans ölçütü olarak önerilmektedir. Bu yeni metrik hem kullanıcı değerlendirmelerini hem de sıralama performansını ölçerek, geleneksel ağırlıklandırılmış ölçütlerle gözden kaçan detayları da dikkate almaktadır. Böylece, sistemin kullanıcı odaklı çıktı performansı daha hassas ve doğru şekilde ölçülebilmektedir.

$$WDCG = \frac{2^{|y-\hat{y}|} \times (2^{\hat{y}} - 1)}{\log_2(\hat{r} + 1)} \tag{3.4}$$

Denklem 3.4'te verilen $WDCG$ değişkeni ağırlıklandırılmış indirgenmiş kümülatif kazanç (Weighted Discounted Cumulative Gain) değerini, y değişkeni ise tariflere ait gerçek değerlendirme puanını ifade etmektedir.

$$iWDCG = \frac{2^{|y_{ideal}-y|} \times (2^{y_{ideal}} - 1)}{\log_2(r_{ideal} + 1)} \tag{3.5}$$

Denklem 3.5'te görülen $iWDCG$ değişkeni ideal ağırlıklandırılmış indirgenmiş kümülatif kazanç (Ideal Weighted Discounted Cumulative Gain) değerini ifade etmektedir.

$$WNDCG = \frac{WDCG}{iWDCG} \tag{3.6}$$

Denklem 3.6'te verilen $WNDCG$ değişkeni ağırlıklandırılmış normalize indirgenmiş kümülatif kazanç (Weighted Normalized Discounted Cumulative Gain) değerini ifade etmektedir.

Ortalama mutlak hata (Mean Absolute Error, MAE), LLM tarafından tahmin edilen ve gerçek değerlendirmeler arasındaki ortalama mutlak hatayı değerlendirmektedir.

$$MAE = \frac{\sum_1^n |y_j - \hat{y}_j|}{n} \tag{3.7}$$

Denklem 3.7'de verilen MAE değişkeni ortalama mutlak hata (Mean Absolute Error) değerini, y_j ve $\hat{y}_j$ değerleri ilgili y ve $\hat{y}$ değişkenlerinin *j*. sıradaki tarife ait sırasıyla gerçek ve LLM tarafından tahmin edilen değerlendirme puanlarını ifade etmektedir. *n* değeri ise toplam değerlendirme adedini ifade etmektedir, LLM tarafından önerilen tarif adedine göre farklı değerler alabilmektedir.

Hata kareler ortalamasının karekökü (Root Mean Squared Error, RMSE), sistem tarafından tahmin edilen tarif değerlendirmeler ile gerçek değerlendirmeler arasındaki hata miktarını ölçmek için kullanılmaktadır.

$$RMSE = \sqrt{\frac{\sum_1^n (y_j - \hat{y}_j)^2}{n}} \tag{3.8}$$

Denklem 3.8'de verilen $RMSE$ hata kareler ortalamasının karekökü değerini ifade etmektedir.

Token kavramı, LLM'ler tarafından işlenen metnin en küçük anlamlı parçalarını ifade eder. Bir token, tam bir kelime, kelime parçacığı veya noktalama işareti gibi küçük metin birimlerini kapsayabilir. Model, gelen metni bu token'lara bölerek işler ve her bir token üzerinden tahmin, analiz veya metin üretimi yapar. Türkçe'de tam bir karşılığı olmamakla beraber "belirteç" veya "jeton" gibi sözcükler ile ifade edilebilse de dilimize de "token" olarak geçmiştir.

Harcanan jeton (token) adetleri, sistem tarafından her bir senaryoda oluşturulan ve LLM'e gönderilen komutlar ile LLM'in ürettiği yanıtlar için kullanılan toplam jeton adedi hesaplanır. Bu hesaplamalar ile farklı komut oluşturma yöntemleri gibi kriterlerin önerilerin masrafı üzerine etkisi değerlendirilmektedir.

LLM yanıt süreleri, her bir senaryo için LLM'in yanıt oluşturma süresi ölçülerek işlem verimliliği analiz edilir. LLM yanıt süreleri gönderilen mesajın uzunluğu, komut içeriği ve görevin karmaşıklık seviyesine göre değişiklik gösterebilmektedir; bu şekilde farklı kriterlerin işlem süresine katkısı değerlendirilebilmektedir.

4. BULGULAR ve TARTIŞMA

Bu bölümde, elde edilen verilerin kapsamlı analizi sonucu ortaya konulan bulgular sunulmuş ve bu sonuçlar mevcut literatür çerçevesinde tartışılmıştır.

4.1. Veri Seti

Önerilen yaklaşımın performansını değerlendirmek amacıyla gerçek bir yemek tarifi sitesinde kullanıcılar tarafından değerlendirilmiş yemek tarifleri ve bu değerlendirmeyi yapan kullanıcılara ait bir veri seti kullanılmış ve algoritma bir simülasyon olarak çalıştırılmıştır.

Deneysel çalışmada, Kaggle web sitesinde yayınlanmış "Food.com Recipes and Interactions" veri seti kullanılmıştır. Bu veri seti, Food.com'daki 18 yıllık kullanıcı değerlendirmeleri ile kullanıcılar tarafından yüklenen yemek tariflerini kapsayan 180.000'den fazla tarif ve 700.000'den fazla değerlendirmeden oluşmaktadır. (Li vd., 2020)

Veri setinden örneklem olarak en çok değerlendirmeyi yapan kullanıcılar ve en çok değerlendirilen yemek tarifleri eşleşmesine uyan 338 kullanıcı ve 2432 yemek tarifi seçilmiştir. Seçilen veriler içerisinde hâlihazırda bir yemek tipi ataması olmadığı için bütün tariflere içerdiği malzemelere göre yemek tipi ataması yapılmıştır.

Seçilen bu veri seti içerisinde ise 0-5 arasındaki değerlendirme puanlarının tamamını en az bir kez kullanmış 10 kullanıcı belirlenmiştir; bu kullanıcılar arasından da 0-5 arasındaki değerlendirme puanlarından her birine eşit miktarlarda puan vermiş olan bir kullanıcı deneysel çalışmada kullanılmak üzere seçilmiştir.

4.2. Uygulama Koşulları

Bu çalışmada, büyük dil modeli (Large Langue Model, LLM), Python programlama dili kullanılarak geliştirilen bir uygulama ile entegre edilmiştir. İlk aşamada, LLM ile uygulama programlama arayüzü (Application Programming Interface, API) entegrasyonu

gerçekleştirilmiştir. OpenAI API entegrasyonu, uygulama ile dil modeli arasında yapılandırılmış veri alışverişi sağlamaktadır.

Kullanıcı sorguları, API sorguları aracılığıyla modele iletilmekte ve alınan yanıtlar, çeşitli veri işleme adımlarından geçirilerek veri tabanına kaydedilmekte ve algoritmada kullanılacak hale getirilmektedir.

Toplamda 288 adet senaryo, beş sefer tekrar edilmiştir. Bu tekrarlarda girdilerin her tekrar için aynı olduğu, bu tez için geliştirilmiş olan bir test senaryosu takip modülü kullanılmıştır. Böylece, bu mekanizma doğru, etkili ve uygulamaya hazır veri akışını mümkün kılmaktadır.

Uygulama içerisinde farklı amaçlar için kullanılan parametreler Çizelge 4.1'de verilmiştir.

Bu parametreler dört ana gruba ayrılmaktadır.

- Komut parametreleri grubu, LLM'e gönderilecek komutun formatını ve içeriğini belirlemek amacıyla kullanılan parametreleri içermektedir.
- Kullanıcının girdiği parametreler grubu, kullanıcılar tarafından arayüz üzerinden girilmesi gereken parametreleri ifade etmektedir. Deneysel çalışmada bu parametreler algoritma tarafından otomatik olarak seçilip atanmaktadır.
- Simülasyon parametreleri grubu, LLM'in performansını değerlendirmek için kullanılan parametrelerdir. Bu parametreler sayesinde kullanıcı etkileşimine gerek kalmadan deneysel koşullarda test yapılabilmektedir.
- LLM parametreleri grubu, kullanılan LLM'in sistem özelliklerine ait parametreleri ifade etmektedir.

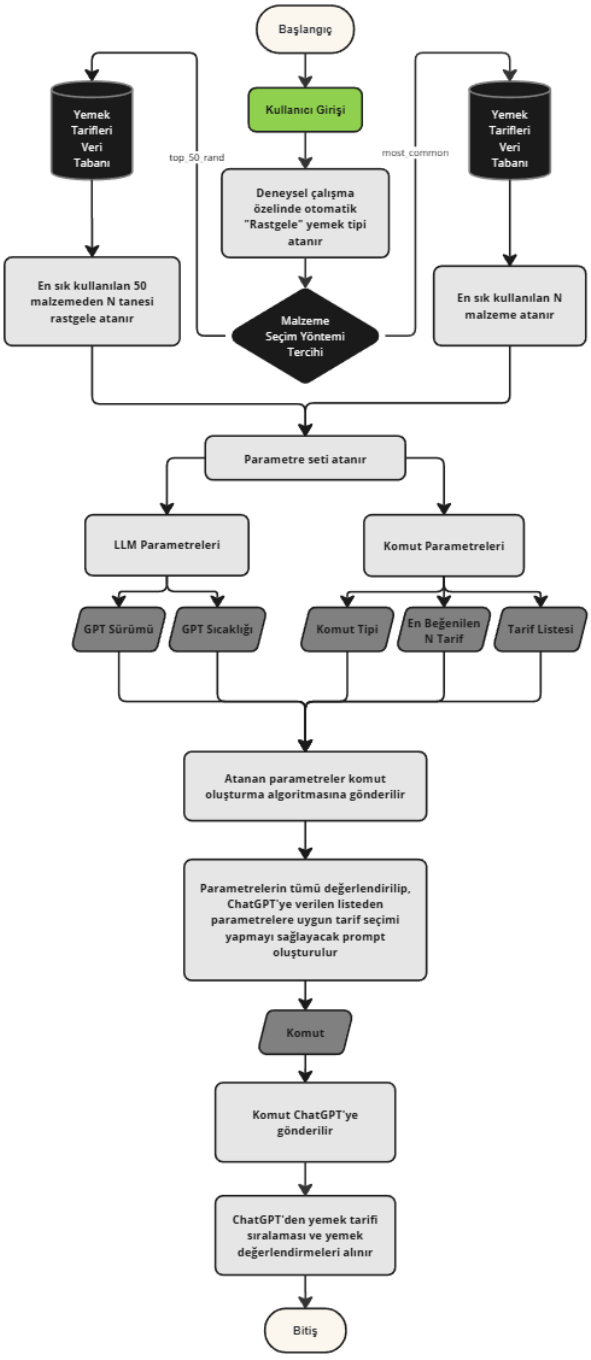
Deneysel çalışma için akış şeması Şekil 4.1'de gösterilmiştir.

Çizelge 4.1. Uygulamaya ait parametreler

Parametre Tipi	Parametre Adı	Açıklama
Komut Parametreleri	Komut tipi (prompt_type)	Performansı etkileyebilecek farklı komut türleri (Few-shot, Zero-shot, Few-shot Role Play, Few-shot Chain of Thought, Zero-shot Role Play, Zero-shot Chain of Thought) analiz edilmiştir.
	Tarif listesi (recipe_list)	Mevcut tariflerin listesini belirtmektedir (sadece bu listeden seçim yapılabilir). Burada verilen liste 20 adet yemek tarifinden oluşmaktadır; bu listenin 10 adedi kullanıcının daha önce yüksek değerlendirme puanı verdiği malzemelerden en az birini içeren tariflerden seçilirken, 10 tanesi ise en düşük değerlendirmeyi almış malzemeleri içeren tariflerden seçilmektedir.
	LLM tarafından önerilmesi beklenen tarif adedi (num_recipes)	LLM tarafından kaç adet tarif önerilmesi beklendiğini belirten parametredir. 1, 5 ve 10 değerlerini alabilmektedir.
Kullanıcının Girdiği Parametreler	Tarif tipi (recipe_type)	Kullanıcının talep ettiği tarif türünü belirtmektedir (örneğin "vegan", "dessert (tatlı)" vb.). Eğer "All (Tüm)" değeri verilirse, "random (rastgele)" olarak dönüştürülmektedir.
	Malzemeler (ingredients)	Kullanıcının öncelik verdiği anahtar malzeme listesini belirtmektedir (örneğin "chicken (tavuk), rice (pirinç), onion (soğan)" vb.). Arayüzden seçilmektedir.
Simülasyon Parametreleri	En beğenilen N tarif (top_n)	Kullanıcının en çok beğendiği tarif sayısının doğruluğa etkileri analiz edilmiştir (3 veya 5). Sadece few-shot tipindeki komutlarda eğitim verisi olarak kullanılmaktadır.
	Malzeme seçimi (ingredient_selection)	Kullanıcının en sık kullandığı malzemeler (most_common) ve rastgele seçilen beş malzeme (top50_rand) kullanılmıştır.
LLM Parametreleri	GPT sıcaklık değeri (gpt_temperature)	Modelin yaratıcılığını ve tutarlılığını etkileyen iki GPT sıcaklık değeri incelenmiştir: 0,2 (daha tutarlı ve az yaratıcı yanıtlar) ve 1,0 (daha yaratıcı ancak düşük tutarlılık).
	GPT sürümü (gpt_version)	GPT-4o ve GPT-4o-mini versiyonlarının performans farklılıkları değerlendirilmiştir.

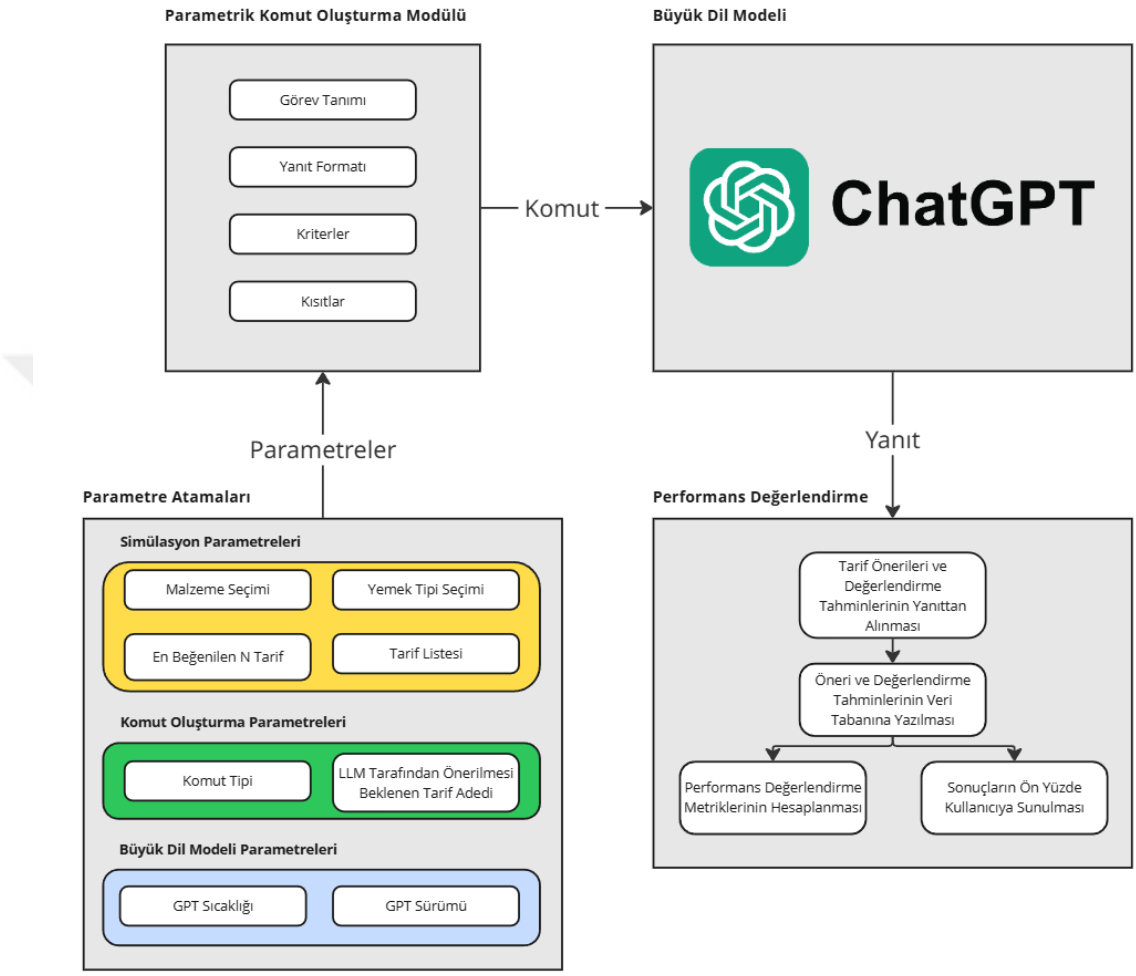
Akış şemasında yeşil ile renklendirilmiş dikdörtgen şeklindeki adımlar kullanıcı tarafından bir seçim yapılarak ya da bir butona tıklanarak gerçekleştirilen interaktif işlemleri; arka plan rengi koyu gri olan paralel kenar şeklindeki adımlar, bir sonraki adımın sonucu olarak ortaya çıkan verileri; arka planı açık gri ile renklendirilmiş adımlar kullanıcı tarafından bir işlem beklenmeden arka planda çalışan algoritmanın otomatik

gerçekleştirdiği işlemleri; arka planı siyah olarak renklendirilmiş eşkenar dörtgenler ise bir karar noktasını ve son olarak yine siyah ile renklendirilmiş silindir prizmalar ise veri tabanına erişimi ifade etmektedir.



Şekil 4.1. Deneysel çalışmanın akış şeması

Uygulama arka planında çalışan modüllerden sonuç olarak ortaya çıkan parametrik komutun LLM'e gönderilmesi ve LLM'den alınan cevabın işlenmesi süreci Şekil 4.2'de görüldüğü şekilde gerçekleşmektedir.



Şekil 4.2. Sistem modülleri ve LLM arasındaki akış şeması

4.3. Deneysel Tasarım

Öneri sistemi üzerinde farklı faktörlerin etkisini anlamak ve sistem performansını optimize etmek amacıyla Deneysel Tasarım (Design of Experiments- DoE) yöntemi kullanılmıştır (Montgomery, 2017). Çalışma kapsamında, birden fazla yanıt değişkeni (WNDCG, NDCG, RMSE, MAE, Yanıt Süresi ve Toplam Jeton Adedi) değerlendirilmiştir. Her bir faktörün ana etkileri ve etkileşimleri analiz edilmiştir.

Yemek tipi seçimi, en yüksek değerlendirme puanlı tarifler ve tarif adedi değişkenleri deneysel çalışmalar süresince sabit değerler olarak algoritmaya eklendiğinden bir faktör olarak değerlendirmeye alınmamıştır.

Deneysel analizde kullanılan faktörler ve çıktılar aşağıdaki gibidir:

- Komut tipi (prompt_type)
- GPT sıcaklığı (gpt_temperature)
- LLM tarafından önerilmesi beklenen tarif adedi (num_recipes)
- En beğenilen N tarif (top_n)
- GPT sürümü (gpt_version)
- Malzeme seçimi (ingredient_selection)

Çalışmada, öneri sisteminin performansını değerlendirmek için aşağıdaki çıktı değişkenleri analiz edilmiştir:

- WNDCG
- NDCG
- RMSE
- MAE
- Yanıt süresi (response_time)
- Toplam jeton adedi (total_tokens)

Yapılan analiz, Python yazılım dilinin 3.10 sürümü ile aşağıdaki kütüphaneler kullanılarak gerçekleştirilmiştir:

- Pandas: Veri işleme ve istatistiksel analizi,
- Matplotlib: Grafiklerin oluşturulması ve görselleştirmesi,
- Openpyxl: Gelişmiş veri görselleştirmesi,
- Numpy: Sayısal hesaplamalar,
- Itertools: Faktör kombinasyonlarının hesaplanması için kullanılmıştır.

4.4. Deneysel Tasarımın Sonuçları

Ana etkiler, her bir faktörün bağımsız olarak yanıt değişkenleri üzerindeki etkilerini ifade eder. Etkileşimler ise iki faktörün birlikte yanıt değişkenleri üzerindeki etkilerini ifade eder.

Şekil 4.3'teki grafiklere göre WNDCG üzerindeki ana etkiler aşağıdaki gibidir.

Şekil 4.3'e göre malzeme seçimi stratejisinde, "en yaygın" yöntemi "rastgele seçilmiş ilk 50" yöntemine göre biraz daha yüksek WNDCG değeri vermiştir. Hata çubuklarının birbirinden ayrışması nedeniyle, bu iki yöntem arasında istatistiksel olarak anlamlı fark olup olmadığını tespit etmek için analiz yapılmalıdır.

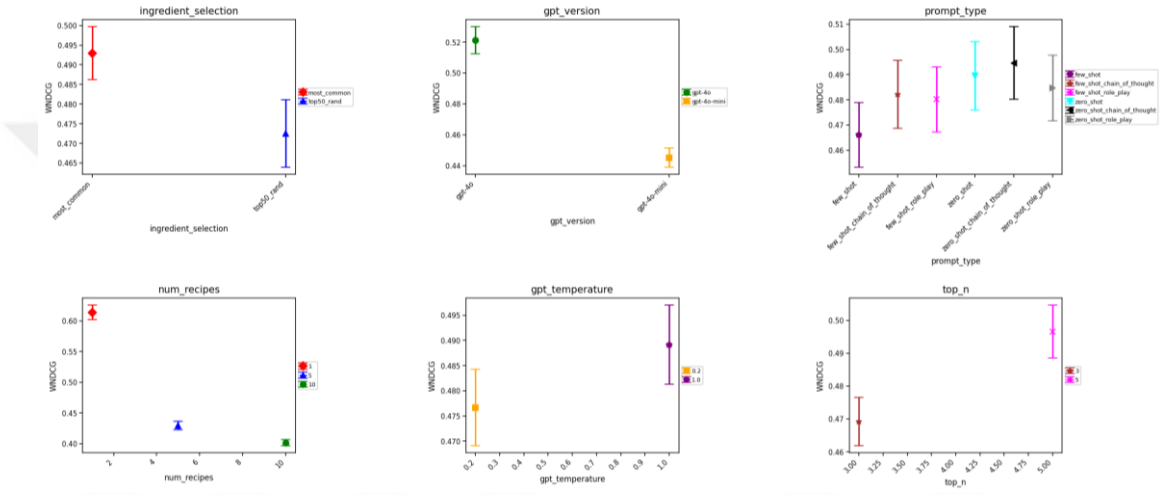
Şekil 4.3'e göre, GPT sürümü açısından "gpt-4o" sürümü, "gpt-4o-mini"ye kıyasla çok daha yüksek WNDCG değeri sunmuştur. "Gpt-4o" kullanımı sıralama performansını belirgin şekilde artırırken, "gpt-4o-mini"da belirgin bir düşüş gözlemlenmiştir. Bu nedenle, GPT sürüm seçimi sistem performansını optimize etmek için kritiktir.

Komut tipi faktöründe, Şekil 4.3'e göre "Sıfır atışlı düşünce zinciri (Zero shot chain of thought)" yöntemi diğerlerine göre daha yüksek WNDCG değeri sağlamıştır. Buna karşılık, "Az atışlı (Few shot)" yöntemi en düşük performansı göstermiştir. Diğer komut tipi türlerinin performansı ise bu iki komut tipinin arasında yer almaktadır. Bu durum, komut tipi seçiminin sıralama performansını küçük fakat fark edilebilir şekilde iyileştirebileceğini göstermektedir.

LLM tarafından önerilen tarif adedi faktöründe, Şekil 4.3'e göre tek tarif (num_recipes=1) kullanımı sıralama performansını dramatik şekilde artırırken, tarif adedinin 5 veya 10 olması WNDCG'de önemli düşüşe yol açmıştır. Bu durum, sistemin tek tarif senaryosunda daha etkili çalıştığını ve tarif adedi artışının performansı olumsuz etkilediğini göstermektedir.

Şekil 4.3'e göre, GPT sıcaklık değeri faktöründe iki farklı sıcaklık arasında anlamlı bir fark bulunamamıştır. Dolayısıyla, incelenen aralıkta GPT sıcaklık değerinin sıralama performansı üzerinde pratik bir etkisi olmadığı belirlenmiştir.

En beğenilen N adet tarif faktöründe ise, Şekil 4.3'e göre 5 değeri az farkla iyi sonuçlar verse de 3 değeriyle arasındaki fark küçüktür ve hata çubukları örtüşmektedir. Bu nedenle, bu faktörün sıralama performansı üzerindeki etkisi önemli değildir.



Şekil 4.3. WNDCG için ana etkiler grafikleri

Şekil 4.4'teki grafiklere göre NDCG üzerindeki ana etkiler aşağıdaki gibidir.

Malzeme seçimi faktöründe, Şekil 4.4'e göre "en yaygın" stratejisinin "rastgele seçilmiş ilk 50" stratejisine göre daha yüksek NDCG değeri verdiği gözlemlenmiştir. Hata çubuklarının örtüşmemesi bu farkın NDCG üzerinde etkisinin olabileceğini düşündürmektedir.

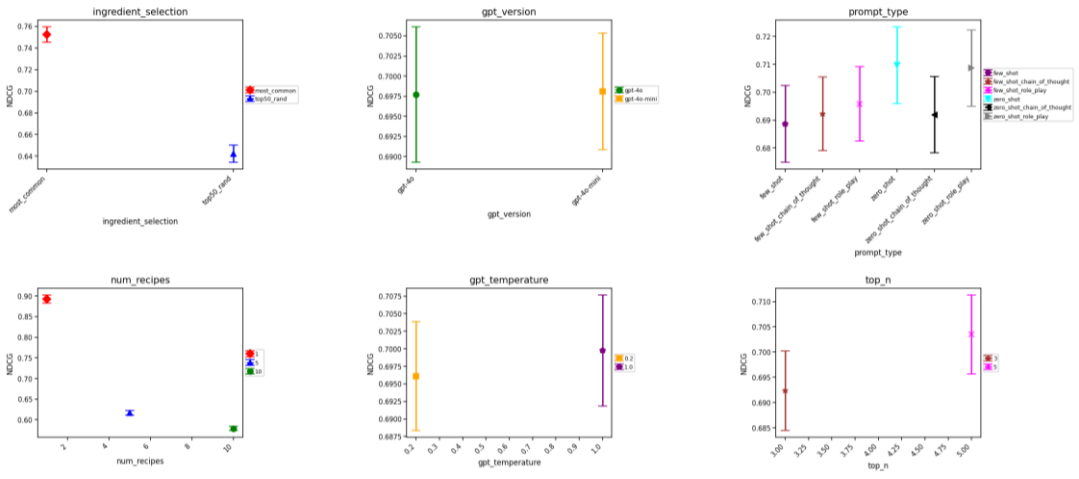
GPT sürümü açısından, Şekil 4.4'e göre "gpt-4o" ile "gpt-4o-mini" arasında anlamlı bir fark bulunamamıştır. "Gpt-4o" hafif olumsuz, "gpt-4o-mini" hafif olumlu etki göstermiştir; ancak farklar küçüktür. Bu nedenle, GPT sürümü seçiminin NDCG üzerinde belirgin bir etkisi yoktur.

Komut tipi açısından, Şekil 4.4'e göre "Sıfır atışlı (Zero shot)" ve "Sıfır atışlı rol yapma (Zero shot role play)" komut tipleri diğerlerine göre daha yüksek NDCG değerleri vermiştir. Özellikle "Sıfır atışlı" en yüksek pozitif etkiyi sağlarken, "Sıfır atışlı düşünce zinciri" ve "Az atışlı düşünce zinciri (Few shot chain of thought)" negatif etki göstermiştir. Diğer komut tiplerinin ise etkileri daha düşüktür. Bu nedenle, komut tipi seçimi NDCG üzerinde belirgin bir etki yapabilmektedir.

LLM tarafından önerilen tarif adedi faktöründe, Şekil 4.4'e göre tek tarif (num_recipes=1) kullanımı NDCG'yi dramatik şekilde artırmıştır. Tarif adedi 1'de NDCG önemli ölçüde yükselirken, 5 ve 10'da belirgin düşüş gözlemlenmiştir. Bu durum, LLM'in önerdiği tarif adedi faktörünün NDCG üzerinde kritik bir etkiye sahip olduğunu göstermektedir.

GPT sıcaklık değeri açısından, Şekil 4.4'e göre 0,2 ve 1,0 arasında belirgin bir fark bulunamamıştır. Her iki seviye de ortalamaya yakın sonuçlar vermiştir. Bu nedenle, GPT sıcaklık değerinin NDCG üzerindeki etkisi ihmal edilebilir düzeydedir.

En beğenilen N adet tarif faktöründe ise, Şekil 4.4'e göre 3 ve 5 değerleri arasında ortalama NDCG'ye göre az fark gözlemlenmektedir. 5 değeri daha yüksek bir sonuç sunmakla birlikte, hata çubuklarının tamamen ayrışmıyor olması, bu farkın çok büyük olmadığını göstermektedir.



Şekil 4.4. NDCG için ana etkiler grafikleri

Şekil 4.5'teki grafiklere göre RMSE üzerindeki ana etkiler aşağıdaki gibidir.

Malzeme seçimi stratejisi faktöründe, Şekil 4.5'e göre "en yaygın" stratejisi, "rastgele seçilmiş ilk 50" stratejisine kıyasla daha yüksek RMSE değeri üretmektedir. "En yaygın" stratejisi RMSE'yi ortalamanın üzerine çıkarırken, "rastgele seçilmiş ilk 50" stratejisi RMSE'yi ortalamanın altında tutmaktadır. Bu durum, "rastgele seçilmiş ilk 50" stratejisinin hata oranını düşürme ve model doğruluğunu artırma potansiyeline sahip olduğu şeklinde değerlendirilebilmektedir.

GPT sürümü faktöründe, Şekil 4.5'e göre "gpt-4o" ile "gpt-4o-mini" arasında belirgin bir fark gözlemlenmektedir. "Gpt-4o" sürümü, daha düşük RMSE değerleri sunarak modelin tahmin doğruluğunu artırırken, "gpt-4o-mini" daha yüksek RMSE değerlerine yol açmaktadır. Bu sonuç, "gpt-4o" sürümünün avantaj sağladığı şeklinde değerlendirilebilmektedir.

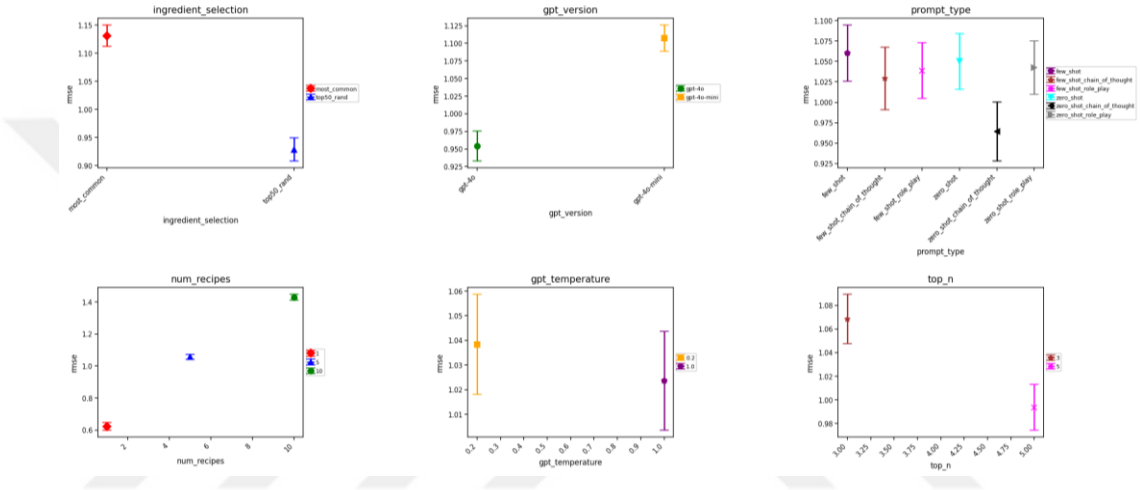
Komut tipi faktöründe, Şekil 4.5'te "Sıfır atışlı düşünce zinciri" komut tipinin diğer türlere göre en düşük RMSE değerini sağladığı görülmektedir. Bu yöntem, hata oranını azaltmada en belirgin etkiye sahipken, "Az atışlı" komut tipi en yüksek RMSE değerine neden olmaktadır. "Sıfır atışlı rol yapma", "Sıfır atışlı" ve "Az atışlı rol yapma" türlerinin etkileri ise orta seviyede seyretmektedir. Bu durum, komut tipi seçiminin modelin tahmin doğruluğu üzerinde önemli bir etki yapabileceği şeklinde değerlendirilebilmektedir.

LLM tarafından önerilmesi beklenen tarif adedi faktöründe, Şekil 4.5'e göre tek tarif (num_recipes=1) kullanımı RMSE değerini dramatik şekilde düşürmektedir. Tarif adedi 1 iken RMSE önemli ölçüde azalırken, 5 ve 10 seviyelerine çıkarıldığında keskin bir artış gözlemlenmektedir; özellikle 10 seviyesinde RMSE en yüksek değere ulaşmaktadır. Bu durum, LLM tarafından önerilen tarif adedi faktörünün RMSE performansı üzerinde belirleyici bir etkiye sahip olduğu şeklinde değerlendirilebilmektedir.

GPT sıcaklık değeri faktöründe ise, Şekil 4.5'e göre 0,2 ve 1,0 seviyeleri arasında belirgin bir fark bulunmamaktadır. Her iki seviye de ortalamaya yakın sonuçlar vermekte ve hata çubukları büyük ölçüde örtüşmektedir. Bu nedenle, GPT sıcaklık değerinin RMSE

performansı üzerinde ihmal edilebilir düzeyde bir etkisi olduğu şeklinde değerlendirilebilmektedir.

En beğenilen N adet tarif faktöründe, Şekil 4.5'e göre, 3 ve 5 değerleri için ortalama RMSE değerleri arasında görece net bir fark gözlemlenmektedir. Hata çubuklarının tamamen ayrışması, bu faktörün RMSE değerinde göz ardı edilemeyecek bir etkiye sahip olabileceğini göstermektedir.



Şekil 4.5. RMSE için ana etkiler grafikleri

Şekil 4.6'teki grafiklere göre MAE üzerindeki ana etkiler aşağıdaki gibidir.

Malzeme seçimi stratejisi incelendiğinde, Şekil 4.6'da "en yaygın" stratejisinin, "rastgele seçilmiş ilk 50" stratejisine göre daha yüksek MAE ürettiği gözlemlenmiştir. "En yaygın" stratejisi MAE'yi ortalamanın üzerine çıkarırken, "rastgele seçilmiş ilk 50" stratejisi MAE'yi ortalamanın altında tutmaktadır. Bu durum, "rastgele seçilmiş ilk 50" stratejisinin hata oranını düşürme ve model doğruluğunu artırma potansiyeline sahip olduğu şeklinde değerlendirilebilmektedir.

GPT sürümü faktörü açısından, Şekil 4.6'da "gpt-4o" ile "gpt-4o-mini" arasında MAE performansı bakımından belirgin bir fark gözlemlenmiştir. "Gpt-4o" sürümü, daha düşük MAE değerleri sunarak modelin tahmin doğruluğunu artırırken, "gpt-4o-mini" daha

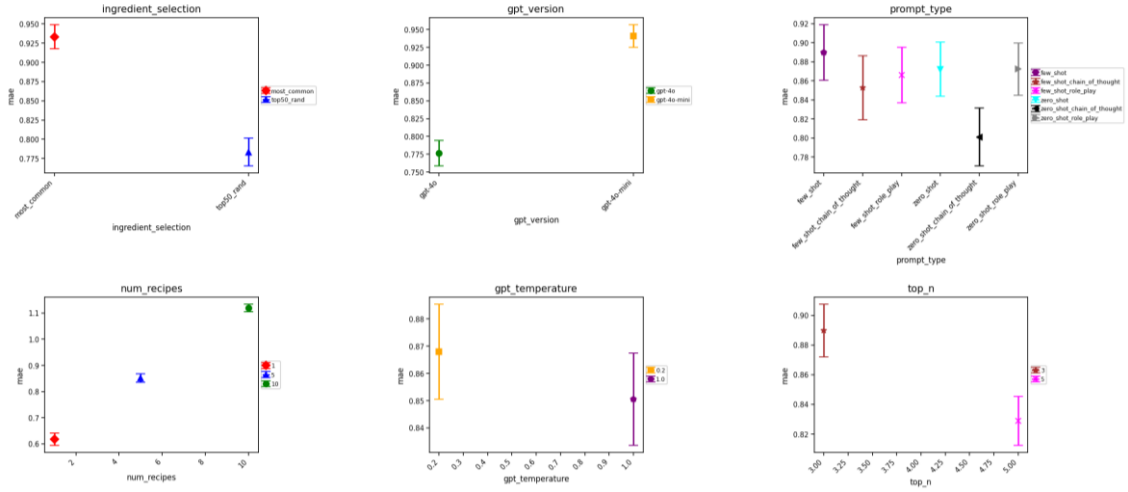
yüksek MAE üretmektedir. Bu sonuç, “gpt-4o” sürümünün modelin tahmin doğruluğunu artırmada avantaj sağladığı şeklinde değerlendirilebilmektedir.

Komut tipi faktörü açısından, Şekil 4.6’da “Sıfır atışlı düşünce zinciri” komut tipinin diğerlerine göre en düşük MAE değerini sağladığı tespit edilmiştir. “Sıfır atışlı düşünce zinciri”, hata oranını düşürmede en belirgin etkiye sahipken, “Az atışlı” komut tipi en yüksek MAE’ye yol açmaktadır. “Sıfır atışlı rol yapma”, “Sıfır atışlı” ve “Az atışlı rol yapma” komut tiplerinin etkileri ise orta seviyede seyretmektedir. Bu bulgular, komut tipi seçiminin modelin tahmin doğruluğu üzerinde önemli bir etki yapabileceği şeklinde değerlendirilebilmektedir.

LLM tarafından önerilmesi beklenen tarif adedi faktörü açısından, Şekil 4.6’da tek tarif (num_recipes=1) kullanımının MAE’yi dramatik şekilde düşürdüğü gözlemlenmiştir. Tarif adedi 1 iken MAE önemli ölçüde azalırken, 5 ve 10 seviyelerine çıkarıldığında keskin bir artış meydana gelmekte; özellikle 10 seviyesinde MAE en yüksek değere ulaşmaktadır. Bu durum, LLM tarafından önerilen tarif adedi faktörünün MAE performansı üzerindeki belirleyici rolünü vurguladığı şeklinde değerlendirilebilmektedir.

GPT sıcaklık değeri faktörü açısından, Şekil 4.6’da 0,2 ve 1,0 seviyeleri arasında MAE bakımından belirgin bir fark bulunamamıştır. Her iki seviye de genel ortalamaya yakın sonuçlar vermekte ve hata çubukları geniş ölçüde örtüşmektedir. Bu nedenle, GPT sıcaklık değerinin bu deneyde incelenen aralıkta MAE performansı üzerinde ihmal edilebilir düzeyde etkiye sahip olduğu şeklinde değerlendirilebilmektedir.

En beğenilen N adet tarif faktörü açısından ise, Şekil 4.6’da 3 ve 5 değeri için ortalama RMSE değerleri arasında belirgin bir farklılık gözlemlenmektedir. Ek olarak hata çubuklarının birbirini kesmemesi nedeniyle bu faktörün RMSE performansı üzerinde kayda değer bir etki yarattığı söylenebilir.



Şekil 4.6. MAE için ana etkiler grafikleri

Şekil 4.7'teki grafiklere göre yanıt süresi üzerindeki ana etkiler aşağıdaki gibidir.

Malzeme seçimi stratejisi incelendiğinde, Şekil 4.7'de "en yaygın" stratejisinin, "rastgele seçilmiş ilk 50" stratejisine göre daha yüksek yanıt süresi ürettiği gözlemlenmiştir. "En yaygın" stratejisi yanıt süresini ortalamanın üzerine çıkarırken, "rastgele seçilmiş ilk 50" stratejisi ortalamanın altında kalmaktadır. Ancak etki büyüklükleri nispeten küçük olduğundan, bu farkın pratik anlamlılığı daha detaylı incelenmelidir.

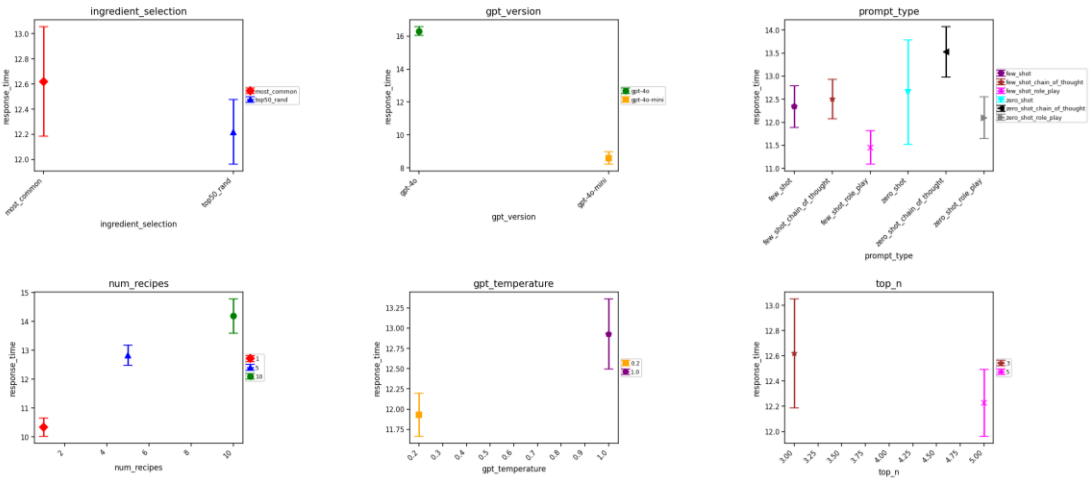
GPT sürümü faktörü açısından, Şekil 4.7'de "gpt-4o" ile "gpt-4o-mini" arasında yanıt süresi bakımından belirgin fark saptanmıştır. "Gpt-4o" daha yüksek yanıt süresi değerleriyle daha yavaş performans sergilerken, "gpt-4o-mini" daha düşük yanıt süresi üretmektedir. Bu durum, "gpt-4o-mini" sürümünün sistem yanıt hızını artırmada avantaj sağladığını göstermektedir.

Komut tipi faktörü incelendiğinde, Şekil 4.7'de "Sıfır atışlı düşünce zinciri" komut tipinin diğerlerine göre en yüksek yanıt süresini ürettiği belirlenmiştir. "Sıfır atışlı düşünce zinciri" yanıt süresini uzatmada belirgin etki gösterirken, "Az atışlı rol yapma" en düşük yanıt süresi değerine sahiptir. Ayrıca, "Sıfır atışlı" komut tipi de nispeten yüksek yanıt süresi sergilerken, "Az atışlı" ve "Sıfır atışlı rol yapma" türlerinin etkileri daha düşük seviyededir. Bu bulgular, komut tipi seçiminin sistem yanıt süresi üzerinde önemli etkiye sahip olduğunu göstermektedir.

LLM tarafından önerilmesi beklenen tarif adedi faktörü incelendiğinde, Şekil 4.7'de tek tarif (num_recipes=1) kullanımının yanıt süresi değerini dramatik şekilde düşürdüğü gözlemlenmiştir. Tarif adedi 1 seviyesinde yanıt süresi önemli ölçüde azalırken, 5 ve 10 seviyelerine çıkarıldığında keskin bir artış meydana gelmekte; özellikle 10 seviyesinde yanıt süresi en yüksek değere ulaşmaktadır. Bu durum, LLM tarafından önerilen tarif adedi faktörünün yanıt süresi performansı üzerindeki belirleyici rolünü teyit etmektedir.

GPT sıcaklık değeri faktörü incelendiğinde, Şekil 4.7'de 0,2 ve 1,0 seviyeleri arasında yanıt süresi bakımından belirgin fark saptanmıştır. 1,0 seviyesi, 0,2'ye göre belirgin şekilde daha yüksek yanıt süresi üretmektedir. Bu durum, GPT sıcaklık değerinin 0,2 olarak ayarlanmasının sistem yanıt süresini kısaltmada avantaj sağladığını göstermektedir.

En beğenilen N adet tarif faktörü incelendiğinde ise, Şekil 4.7'de 3 ve 5 seviyeleri arasında yanıt süresi açısından küçük bir fark gözlemlenmiştir. 3 değeri negatif etkiyle yanıt süresini artırırken, 5 değeri pozitif etkiyle yanıt süresini azaltmaktadır. Ancak, bu etkiler nispeten küçük olup hata çubuklarının örtüşmesi, bu faktörün yanıt süresi üzerinde sınırlı bir etkiye sahip olduğunu göstermektedir.



Şekil 4.7. Yanıt süresi için ana etkiler grafikleri

Şekil 4.8'deki grafiklere göre harcanan jeton adedi üzerindeki ana etkiler aşağıdaki gibidir.

Malzeme seçimi stratejileri incelendiğinde, Şekil 4.8'de "rastgele seçilmiş ilk 50" yöntemi, "en yaygın" yöntemine kıyasla toplam jeton sayısını (total_tokens) büyük oranda artırmaktadır. Bu bulgu, "rastgele seçilmiş ilk 50" yöntemiyle modelin daha fazla jeton kullandığını ve dolayısıyla çıktı uzunluğunun ve detay seviyesinin artabileceğini göstermektedir.

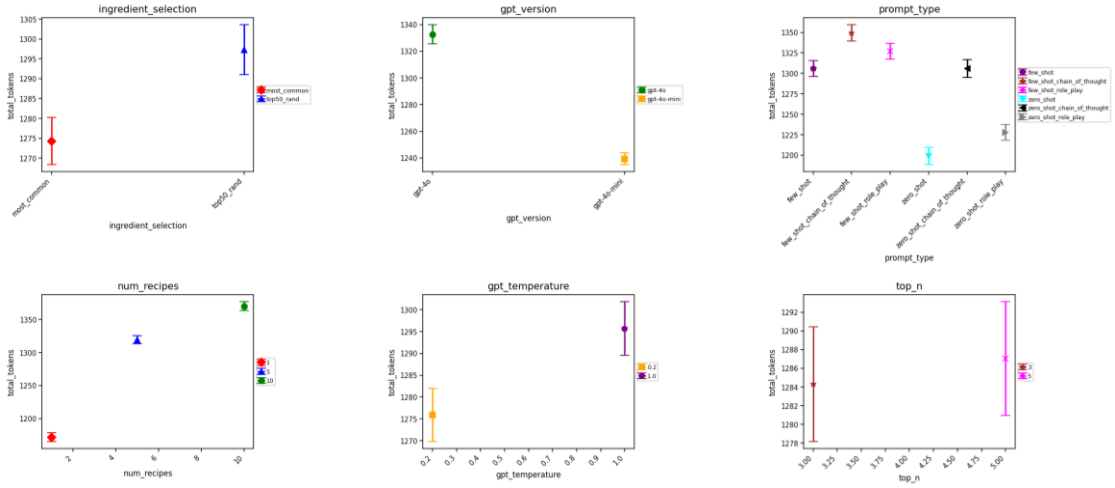
GPT sürümü incelendiğinde, "gpt-4o" sürümünün, "gpt-4o-mini"ye göre toplam jeton sayısını belirgin ve büyük oranda artırdığı tespit edilmiştir. Bu durum, "gpt-4o" sürümünün aynı görev için daha fazla jeton kullanmasına yol açtığını ve modelin çıktı uzunluğu ile bilgi içeriğini etkileyebileceğini göstermektedir.

Komut tipi faktörü değerlendirildiğinde, "Sıfır atışlı" ve "Sıfır atışlı rol yapma" komut tiplerinin diğerlerine göre daha düşük jeton kullanımı ile sonuçlandığı gözlemlenmiştir. Buna karşılık, "Az atışlı düşünce zinciri" komut tipinin modelin belirgin şekilde daha fazla jeton kullanmasına neden olduğu saptanmıştır. Bu bulgu, komut tipi seçiminin model çıktılarının uzunluğu ve potansiyel bilgi yoğunluğu üzerinde etkili olduğunu göstermektedir.

LLM tarafından önerilmesi beklenen tarif adedi faktörü incelendiğinde, tek tarif (num_recipes=1) kullanımının jeton kullanımını önemli ölçüde azalttığı, tarif adedinin artması durumunda (örneğin 5 ve 10) ise jeton kullanımının belirgin şekilde yükseldiği tespit edilmiştir. Bu sonuç, LLM tarafından önerilen tarif adedinin model çıktısının uzunluğu ve bilgi içeriği üzerinde güçlü ve doğru orantılı bir etkiye sahip olduğunu ortaya koymaktadır.

GPT sıcaklık değeri incelendiğinde, sıcaklık değerinin 0.2'den 1.0'a yükseltilmesinin modelin ürettiği toplam jeton sayısını artırdığı saptanmıştır. Bu durum, sıcaklık ayarının model çıktılarının uzunluğu ve çeşitliliği üzerinde etkili olduğunu düşündürmektedir.

Kullanıcının beğendiği tarif sayısının (3 veya 5) toplam jeton sayısı üzerindeki etkisi incelendiğinde ise, bu faktörün model çıktısının uzunluğu üzerinde büyük bir etki yaratmadığı tespit edilmiştir.



Şekil 4.8. Harcanan jeton adedi için ana etkiler grafikleri

WNDCG için etkileşim grafikleri Şekil 4.9'da verilmiştir, diğer çıktı değişkenleri için etkileşim grafikleri Ek 1-5'te yer almaktadır. Bu grafiklerden elde edilen sonuçlar aşağıda özetlenmiştir.

Şekil 4.9'da görülen WNDCG etkileşim grafikleri incelendiğinde, faktörler arasında önemli karşılıklı etkileşimler olduğu görülmektedir. Özellikle GPT sürümü (gpt-4o ve gpt-4o-mini) ile prompt tipi arasında güçlü bir etkileşim vardır; gpt-4o daha iyi performans gösterse de uygun prompt tipi kullanılan sürüme göre değişmektedir. Benzer şekilde, prompt tipinin etkisi LLM tarafından önerilmesi beklenen tarif adedine, GPT sıcaklığına ve en beğenilen N adet tarif sayısına göre de önemli ölçüde farklılaşmaktadır. Malzeme seçimi (most_common ve top50_rand) açısından, top50_rand seçimi genellikle most_common seçimine göre daha yüksek WNDCG sağlarken, bu seçimin avantajı kullanılan GPT sürümüne bağlıdır ve özellikle GPT sıcaklığı arttıkça performansı most_common seçimine göre daha fazla düşme eğilimindedir. GPT sıcaklığı da prompt tipi, LLM tarafından önerilmesi beklenen tarif adedi ve en beğenilen N adet tarif sayısı gibi faktörlerle etkileşime girerken, en beğenilen N adet tarif sayısının artması özellikle yüksek sıcaklıklarda veya LLM'den az sayıda tarif beklenmesi durumunda performansı düşürme eğilimindedir. GPT sürümünün LLM tarafından önerilmesi beklenen tarif adedi ile etkileşimi ise daha zayıf görünmektedir.

Ek 1'de NDCG metrikleri için sunulan etkileşim grafikleri incelendiğinde, faktörler arasında yine önemli ve karmaşık ilişkiler olduğu görülmektedir. GPT sürümü (gpt-4o ve gpt-4o-mini) ile prompt tipi arasındaki güçlü etkileşim burada da belirgindir; gpt-4o daha yüksek NDCG sağlasa da hangi prompt tipinin en iyi performansı vereceği kullanılan GPT sürümüne göre değişmektedir. Prompt tipinin etkisi; LLM tarafından önerilmesi beklenen tarif adedine, GPT sıcaklığına ve en beğenilen N adet tarif sayısına göre de farklılaşmaktadır. Malzeme seçimi (most_common ve top50_rand) açısından, top50_rand seçimi, özellikle gpt-4o sürümü ile kullanıldığında most_common seçimine göre iyi sonuç verirken, GPT sıcaklığı arttıkça performansı most_common seçimine göre daha keskin bir şekilde düşmektedir. Bu durum, malzeme seçimi ile hem GPT sürümü hem de sıcaklık arasında önemli etkileşimler olduğunu gösterir. En beğenilen N adet tarif sayısının artırılmasının etkisi ise özellikle LLM tarafından önerilmesi beklenen tarif adedi düşük olduğunda olumsuz olmaktadır. Ayrıca, bu etkinin kullanılan prompt tipine ve GPT sürümüne göre de değiştiği görülmektedir. GPT sürümünün LLM tarafından önerilmesi beklenen tarif adedi veya GPT sıcaklığı ile etkileşimleri ise diğerlerine kıyasla daha zayıf görünmektedir.

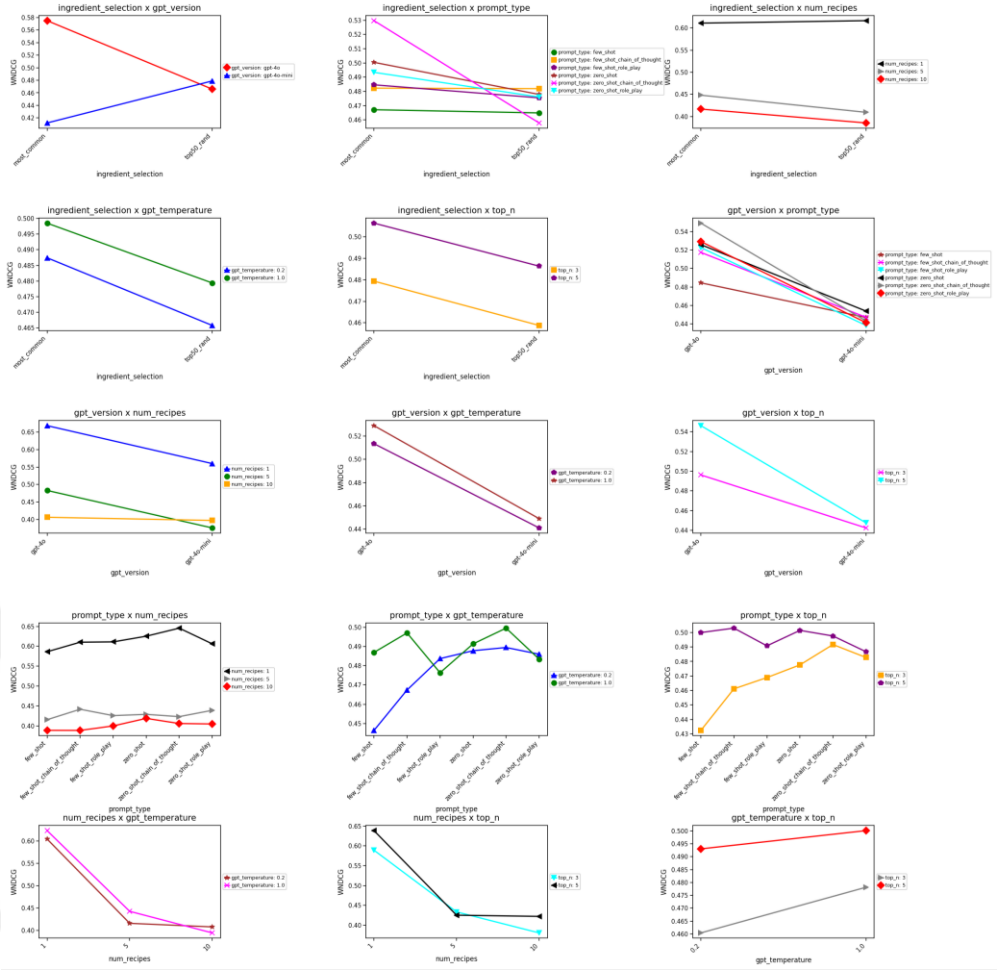
Ek 2'de. GPT sürümü (gpt-4o ve gpt-4o-mini) ile prompt tipi arasındaki güçlü etkileşim RMSE'de de gözlenmektedir; gpt-4o daha düşük hata (daha iyi performans) sağlasa da en düşük hatayı veren prompt tipi kullanılan GPT sürümüne göre değişmektedir. İlginç bir şekilde, GPT sürümü ile GPT sıcaklığı arasında da belirgin bir etkileşim vardır: gpt-4o düşük sıcaklıkta daha az hata yaparken (daha iyi), sıcaklık arttıkça hatası gpt-4o-mini sürümüne göre daha hızlı artmaktadır, hatta yüksek sıcaklıklarda gpt-4o-mini daha düşük hata verebilmektedir. Prompt tipinin hataya etkisi; LLM tarafından önerilmesi beklenen tarif adedine, GPT sıcaklığına ve en beğenilen N adet tarif sayısına göre kuvvetli bir şekilde değişmektedir. Malzeme seçimi (most_common ve top50_rand) açısından, top50_rand seçimi, gpt-4o sürümü ile kullanıldığında belirgin şekilde daha düşük hata verirken, most_common seçimi gpt-4o-mini sürümü ile (daha düşük RMSE) vermektedir. Sıcaklık arttıkça top50_rand seçiminin hatası most_common seçimine göre daha hızlı artmaktadır. Hem sıcaklığın artması hem de en beğenilen N adet tarif sayısının artması, özellikle LLM tarafından önerilmesi beklenen tarif adedi düşük olduğunda, hatayı (RMSE) daha belirgin bir şekilde artırmaktadır (performansı düşürmektedir).

Ek 3'te verilen MAE için etkileşim grafikleri incelendiğinde, RMSE grafikleriyle benzer etkileşimler gözlemlenmektedir. GPT sürümü ile prompt tipi arasında yine güçlü bir etkileşim mevcuttur; gpt-4o daha düşük MAE (daha iyi performans) sağlasa da en düşük hatayı veren prompt tipi kullanılan GPT sürümüne bağlı olarak değişmektedir. GPT sürümü ile GPT sıcaklığı arasındaki etkileşim de RMSE'deki ile aynıdır: gpt-4o düşük sıcaklıkta daha az hata yaparken (daha iyi performans), sıcaklık arttıkça hatası gpt-4o-mini sürümüne göre daha hızlı artmakta ve yüksek sıcaklıklarda gpt-4o-mini daha avantajlı (daha düşük MAE) hale gelebilmektedir. Prompt tipinin MAE üzerindeki etkisi; LLM tarafından önerilmesi beklenen tarif adedine, GPT sıcaklığına ve en beğenilen N adet tarif sayısına göre kuvvetli bir şekilde farklılaşmaktadır. Malzeme seçimi (most_common ve top50_rand) açısından bakıldığında, top50_rand seçimi gpt-4o sürümü ile kullanıldığında belirgin şekilde daha düşük hata verirken, most_common seçimi gpt-4o-mini sürümü ile daha iyi (daha düşük MAE) sonuçlanmaktadır. Ayrıca, sıcaklık arttıkça top50_rand seçiminin hatası most_common seçimine göre daha hızlı yükselmektedir. Hem GPT sıcaklığının artması hem de en beğenilen N adet tarif sayısının artması, özellikle LLM tarafından önerilmesi beklenen tarif adedi az olduğunda, MAE'yi belirgin bir şekilde artırmakta, bir başka deyişle performansı düşürmektedir.

Ek 4'te paylaşılan yanıt süresi için etkileşim grafikleri incelendiğinde, bazı dikkat çekici etkileşimler görülmektedir. GPT sürümü (gpt-4o ve gpt-4o-mini) ile prompt tipi arasında güçlü bir etkileşim vardır; gpt-4o hemen hemen tüm durumlarda gpt-4o-mini sürümünden belirgin şekilde daha hızlıdır (daha düşük yanıt süresi), ancak en hızlı prompt tipi ve hız farkı kullanılan sürüme ve prompt'a göre değişebilmektedir. GPT sürümleri arasındaki bu hız farkı, LLM tarafından önerilmesi beklenen tarif adedi arttıkça daha da belirginleşmektedir; gpt-4o daha fazla tarif istendiğinde gpt-4o-mini'ye kıyasla daha az yavaşlamaktadır. Malzeme seçimi (most_common ve top50_rand) de GPT sürümü ile etkileşime girmektedir: top50_rand seçimi gpt-4o ile daha hızlı yanıt verirken, most_common seçimi gpt-4o-mini ile daha hızlıdır. Ayrıca, top50_rand seçiminin yanıt süresi, GPT sıcaklığı veya en beğenilen N adet tarif sayısı arttıkça most_common seçimine göre artmakta, yani daha çok yavaşlamaktadır. Prompt tipinin yanıt süresine etkisi; LLM tarafından önerilmesi beklenen tarif adedine, GPT sıcaklığına ve en beğenilen N adet tarif sayısına göre kuvvetli ve karmaşık bir şekilde değişmektedir.

Dikkat çekici bir diğer nokta ise, GPT sıcaklığının veya en beğenilen N adet tarif sayısının yanıt süresini artırma (yavaşlatma) etkisinin, LLM tarafından önerilmesi beklenen tarif adedi fazla olduğunda daha belirgin olmasıdır; bu durum RMSE ve MAE metriklerindeki eğilimin tersidir. GPT sürümünün sıcaklık veya en beğenilen N adet tarif sayısı ile belirgin bir etkileşimi gözlenmese de gpt-4o'nun genel olarak çok daha hızlı olduğu görülmektedir.

Ek 5'te verilen toplam jeton sayısı için etkileşim grafikleri incelendiğinde, bu metriğin eğilimlerinin ve etkileşimlerinin yanıt süresi (response_time) ile oldukça benzer olduğu görülmektedir. GPT sürümü (gpt-4o ve gpt-4o-mini) ile prompt tipi arasında güçlü bir etkileşim vardır; gpt-4o genellikle daha az jeton kullansa da (daha verimli), en verimli (en az jeton harcayan) prompt tipi kullanılan GPT sürümüne ve spesifik prompt'a göre değişmektedir. İki GPT sürümü arasındaki jeton kullanım farkı, LLM tarafından önerilmesi beklenen tarif adedi arttıkça açılmaktadır; gpt-4o daha fazla tarif için gpt-4o-mini'ye kıyasla daha az ek jeton harcamaktadır. Malzeme seçimi (most_common ve top50_rand) de GPT sürümü ile etkileşime girmektedir: top50_rand seçimi gpt-4o ile daha az jeton kullanırken, most_common seçimi gpt-4o-mini ile daha verimlidir. Buna ek olarak, top50_rand seçiminin jeton kullanımı, GPT sıcaklığı veya en beğenilen N adet tarif sayısı arttıkça most_common seçimine göre daha fazla artmaktadır. Prompt tipinin jeton kullanımına etkisi; LLM tarafından önerilmesi beklenen tarif adedine, GPT sıcaklığına ve en beğenilen N adet tarif sayısına göre kuvvetli ve karmaşık bir şekilde değişmektedir. Yanıt süresinde olduğu gibi, GPT sıcaklığının veya en beğenilen N adet tarif sayısının jeton kullanımını artırma etkisi, LLM tarafından önerilmesi beklenen tarif adedi fazla olduğunda daha belirgindir. GPT sürümünün sıcaklık veya en beğenilen N adet tarif sayısı ile belirgin bir etkileşimi gözlenmemekle birlikte, gpt-4o'nun genel olarak daha az jeton kullandığı (daha verimli olduğu) söylenebilmektedir.



Şekil 4.9. WNDCG için etkileşim grafikleri

4.5. İstatistiksel Analiz

Bu bölümde bağımsız değişkenlerin bağımlı değişkenler üzerindeki etkisi istatistiksel olarak değerlendirilmiştir

İstatistiksel testler, parametrik ve parametrik olmayan olmak üzere iki ana kategoriye ayrılmaktadır. Parametrik testler; verilerin normal dağılım, varyans homojenliği gibi varsayımları karşıladığı durumlarda kullanılmaktadır ve varsayımlar sağlandığında güçlü ve güvenilir sonuçlar vermektedirler. Parametrik olmayan testler ise daha esnek olup, varsayımların sağlanmadığı durumlarda da kullanılabilmektedirler.

Bu tezde kullanılacak yöntemin belirlenmesi için öncelikle parametrik testlerdeki varsayımların sağlanıp sağlanmadığı kontrol edilmiştir. Ek 6, Ek 7 ve Ek 8'de sunulan sonuçlara göre normallik varsayımının reddedildiği ve kovaryans homojenliğinin ise sağlanmadığı görülmektedir.

Bu sebeple, parametrik olmayan testler ile devam edilmesi kararlaştırılmıştır. Bağımsız değişkenlerin bağımlı değişkenler üzerindeki etkilerini saptamak için parametrik olmayan yöntemlerden Kruskal-Wallis, Epsilon-Kare ve permütasyonel varyans analizi (PERMANOVA) testleri kullanılmıştır.

4.5.1. Kruskal-Wallis ve Epsilon Kare Testleri

Bağımsız değişkenlerin bağımlı değişkenler üzerindeki etkilerini saptamak için Kruskal-Wallis parametrik olmayan testi (Siegel ve Castellan, 1988) uygulanmıştır. Bu test, Minitab Statistical Software programı kullanılarak uygulanmıştır. Anlamlılık düzeyi olarak $\alpha = 0,05$ seçilmiştir.

Kruskal-Wallis testindeki H değeri, gruplar arası varyansı temsil eden bir test istatistiğidir ve grupların sıralama ortalamaları arasında istatistiksel olarak anlamlı bir fark olup olmadığını göstermektedir. Yüksek bir H değeri, gruplar arasında daha büyük bir fark olduğunu işaret etmektedir (Kruskal ve Wallis, 1952).

Bu testin tercih edilme nedeni, deney sonuçlarında ANOVA ve MANOVA gibi parametrik testlerin gerektirdiği normallik ve kovaryans matrisi eşitliği varsayımlarının karşılanmamasıdır (Hollander vd., 2013; Büyüköztürk vd., 2018). Kruskal-Wallis parametrik olmayan testi, söz konusu varsayımlara gereksinim duymadan uygulanabilmesi avantajını sunmaktadır (Özdamar, 2019).

İstatistiksel olarak anlamlı fark saptanan durumlar için ise, etki büyüklüğünü belirlemek amacıyla Epsilon Kare (ε^2) istatistiği hesaplanmıştır. Epsilon Kare (ε^2), Kruskal-Wallis gibi parametrik olmayan testlerde istatistiksel olarak anlamlı farklılıkların pratik önemini değerlendiren bir etki büyüklüğü ölçüsüdür (Field, 2018). Bu istatistik, grup

farklılıklarının varyansın ne kadarını açıkladığını göstererek, bulguların istatistiksel olarak anlamlılığını yorumlamaya yardımcı olur.

Kruskal-Wallis testi, her bir bağımsız değişkenin, bağımlı değişkenler üzerindeki etkilerini ortaya koymuştur (Çizelge 4.2). Tablolarda yer alan df (degrees of freedom) ilgili faktörün serbestlik derecesini ifade etmektedir.

Çizelge 4.2'deki sonuçlara göre bağımsız değişkenlerin bağımlı değişkenlere olan etkisi aşağıda verilmiştir.

WNDCG üzerinde, GPT sürümü, malzeme seçimi, en beğenilen N adet tarif ve LLM tarafından önerilmesi beklenen tarif adedi istatistiksel olarak anlamlı etki göstermektedir. Bu durum, WNDCG'nin çeşitli faktörler tarafından belirlendiğini ortaya koyarken, komut tipi ve GPT sıcaklık değeri istatistiksel olarak anlamlı etki göstermemekte, dolayısıyla sıralama kalitesinde doğrudan rol oynamamaktadır.

NDCG açısından, malzeme seçimi ve LLM tarafından önerilmesi beklenen tarif adedi önemli etki yapmaktadır. Buna karşılık, GPT sürümü, komut tipi, GPT sıcaklık değeri ve en beğenilen N adet tarif üzerinde belirgin bir etki tespit edilememiştir.

RMSE'de, GPT sürümü, malzeme seçimi ve LLM tarafından önerilen tarif adedi istatistiksel olarak anlamlı farklılıklar yaratırken; komut tipi, GPT sıcaklık değeri ve en beğenilen N adet tarifin etkisi belirgin bulunmamaktadır.

MAE'de, GPT sürümü, komut tipi, malzeme seçimi ve LLM tarafından önerilen tarif adedi istatistiksel olarak anlamlı etki gösterirken; GPT sıcaklık değeri ve en beğenilen N adet tarifin MAE üzerinde belirgin etkisi olmadığı saptanmıştır.

Yanıt süresi üzerinde, GPT sürümü, komut tipi, GPT sıcaklık değeri ve LLM tarafından önerilen tarif adedi istatistiksel olarak anlamlı etki sergilerken; malzeme seçimi ve en beğenilen N adet tarif belirgin rol oynamamaktadır.

Kaynak kullanımı (toplam jeton adedi) açısından, malzeme seçimi, GPT sıcaklık değeri ve LLM tarafından önerilen tarif adedi istatistiksel olarak anlamlı etkilere sahipken; GPT sürümü, komut tipi ve en beğenilen N adet tarif faktörleri toplam jeton adedi üzerinde belirgin bir farklılık yaratmamaktadır.

Çizelge 4.2. Kruskal-Wallis testi sonuçları (p-degeri ≤ 0,05 ise koyu olarak işaretlenmiştir)

Bağımlı Değişken	Etki	df	H-değeri	p-değeri
WNDCG	Komut Tipi	5	2,73	0.741
	GPT Sürümü	1	54,44	**<0,001***
	Malzeme Seçimi	1	13,47	**<0,001***
	GPT Sıcaklık Değeri	1	1,11	0.291
	En Beğenilen N Adet Tarif	1	4,02	**0.045***
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	2	220,32	**<0,001***
NDCG	Komut Tipi	5	2,02	0.847
	GPT Sürümü	1	0,33	0.563
	Malzeme Seçimi	1	114,96	**<0,001***
	GPT Sıcaklık Değeri	1	0,25	0.616
	En Beğenilen N Adet Tarif	1	2,96	0.085
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	2	387,90	**<0,001***
RMSE	Komut Tipi	5	3,19	0.670
	GPT Sürümü	1	42,85	**<0,001***
	Malzeme Seçimi	1	47,69	**<0,001***
	GPT Sıcaklık Değeri	1	0,92	0.337
	En Beğenilen N Adet Tarif	1	2,47	0.116
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	2	551,07	**<0,001***
MAE	Komut Tipi	5	5,18	0.394
	GPT Sürümü	1	53,26	**<0,001***
	Malzeme Seçimi	1	44,69	**<0,001***
	GPT Sıcaklık Değeri	1	1,28	0.258
	En Beğenilen N Adet Tarif	1	3,76	0.053
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	2	249,28	**<0,001***
Yanıt Süresi	Komut Tipi	5	11,46	**0.043***
	GPT Sürümü	1	697,71	**<0,001***
	Malzeme Seçimi	1	0,04	0.834
	GPT Sıcaklık Değeri	1	13,29	**<0,001***
	En Beğenilen N Adet Tarif	1	0,10	0.749
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	2	96,54	**<0,001***
Toplam Jeton Adedi	Komut Tipi	5	159,15	**<0,001***
	GPT Sürümü	1	95,20	**<0,001***
	Malzeme Seçimi	1	8,74	**<0,001***
	GPT Sıcaklık Değeri	1	6,77	**0.009***
	En Beğenilen N Adet Tarif	1	0,21	0.646
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	2	473,46	**<0,001***

Kruskal-Wallis testi ile istatistiksel olarak anlamlı bulunan etkilerin büyüklükleri, Epsilon Kare (ε^2) değerleri ile incelenmiştir (Çizelge 4.3).

Çizelge 4.3'te görülen değerlere göre yapılan bu analiz, teste dahil edilmiş her bir bağımlı değişken için aşağıdaki sonuçları vermiştir.

WNDCG üzerinde en belirgin etki, LLM tarafından önerilmesi beklenen tarif adedinden kaynaklanmaktadır; bu etkinin büyük (ε^2=0,162) olduğu söylenebilir. Buna karşın, GPT sürümü (ε^2=0,040), malzeme seçimi (ε^2=0,009) ve kullanıcının beğendiği tarif sayısının (ε^2=0,002) etkileri küçüktür.

NDCG için de baskın unsur, LLM tarafından önerilmesi beklenen tarif adedidir (ε^2=0,286); etkisi büyük kabul edilirken, malzeme seçiminin katkısı (ε^2=0,084) orta düzeydedir. Bu durum, NDCG açısından LLM tarafından önerilmesi beklenen tarif adedinin belirleyici rol oynadığını ortaya koymaktadır

RMSE'de en büyük etki yine LLM tarafından önerilmesi beklenen tarif adedine (ε^2=0,407) ait olup, bu etki büyüktür; GPT sürümü (ε^2=0,031) ve malzeme seçiminin (ε^2=0,035) etkileri ise küçüktür. Bu durum, RMSE'nin azaltılmasında LLM tarafından önerilmesi beklenen tarif adedinin artırılmasının kritik bir öneme sahip olduğunu işaret eder.

MAE açısından, en belirgin etki LLM tarafından önerilmesi beklenen tarif adedinden (ε^2=0,183) gelmekte ve etkisi büyük düzeydedir; GPT sürümü (ε^2=0,039) ve malzeme seçiminin (ε^2=0,032) katkıları ise küçüktür. Yani, MAE'yi düşürmede LLM tarafından önerilmesi beklenen tarif adedinin önemli rolü bulunurken, diğer faktörlerin de küçük katkılar sağladığı gözlemlenmektedir.

Yanıt süresinde ise, LLM tarafından önerilmesi beklenen tarif adedinin (ε^2=0,070) orta düzeyde, komut tipinin (ε^2=0,005) ise küçük düzeyde etkisi mevcutken, GPT sıcaklık değerinin (ε^2=0,009) etkisi de küçüktür. Bu durum, yanıt süresinin optimize edilmesinde birden fazla faktörün rol oynadığını göstermektedir.

Son olarak, toplam jeton adedi üzerinde en büyük etki LLM tarafından önerilmesi beklenen tarif adedine (ε^2=0,349) ait olup, etkisi büyüktür; komut tipi (ε^2=0,114) ve GPT

sürümünün (ε^2=0,070) etkileri orta, malzeme seçimi (ε^2=0,006) ve GPT sıcaklık değerinin (ε^2=0,004) etkileri ise küçüktür. Bu, kaynak kullanımında temel belirleyici unsurun LLM tarafından önerilmesi beklenen tarif adedi olduğunu ortaya koymaktadır.

Sonuçlara göre, girdi olarak LLM tarafından önerilmesi beklenen tarif adedinin, öneri sisteminin performansı üzerinde en belirleyici faktör olduğu görülmektedir. Daha fazla veri ile eğitilen ÜYZ modelleri, hem sıralama kalitesi (WNDCG, NDCG) hem de tahmin doğruluğu (RMSE, MAE) açısından daha başarılı öneriler sunmaktadır. Bu durum, makine öğrenimi literatüründeki, model performansının genellikle veri miktarıyla doğru orantılı olduğu yönündeki genel görüşle uyumludur (Hastie vd., 2009).

Özetle ÜYZ tabanlı yemek tarifi öneri sistemlerinin başarısı, çok sayıda faktörün etkileşimine bağlıdır. Kruskal-Wallis ve Epsilon-Kare test sonuçlarına göre bu faktörler arasında en önemlisi, girdi olarak LLM tarafından önerilmesi beklenen tarif adedidir. Diğer faktörler (GPT sürümü, malzeme seçimi, komut tipi, GPT sıcaklık değeri) ise daha düşük düzeyde etkiye sahiptir.

Çizelge 4.3. Epsilon-Kare testi sonuçları

Bağımlı Değişken	Etki	H	ε^2
WNDCG	GPT Sürümü	54,440	0,040
	Malzeme Seçimi	13,470	0,009
	En Beğenilen N Adet Tarif	4,020	0,002
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	220,320	0,162
NDCG	Malzeme Seçimi	114,960	0,084
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	387,900	0,286
RMSE	GPT Sürümü	42,850	0,031
	Malzeme Seçimi	47,690	0,035
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	551,070	0,407
MAE	GPT Sürümü	53,260	0,039
	Malzeme Seçimi	44,690	0,032
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	249,280	0,183
Yanıt Süresi	Komut Tipi	11,460	0,005
	GPT Sıcaklık Değeri	13,290	0,009
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	96,540	0,070
Toplam Jeton Adedi	Komut Tipi	159,150	0,114
	GPT Sürümü	95,200	0,070
	Malzeme Seçimi	8,740	0,006
	GPT Sıcaklık Değeri	6,770	0,004
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	473,460	0,349

4.5.2. Permütasyonel Varyans Analizi

Permütasyonel varyans analizi (PERMANOVA), bağımsız değişkenlerin farklı seviyeleri arasında, bağımlı değişkenlerin dağılımında istatistiksel olarak anlamlı bir fark olup olmadığını, mesafe tabanlı bir yaklaşımla test etmektedir (Anderson, 2001; McArdle ve Anderson, 2001).

Deneysel çalışmada kullanılan veri setinin çok değişkenli normallik varsayımını karşılamaması ve parametrik çok değişkenli varyans analizi (MANOVA) varsayımlarına uygun olmaması nedeniyle PERMANOVA tercih edilmiştir.

Bu analiz, R istatistiksel programlama dili ve "vegan" kütüphanesi (Oksanen vd., 2020) kullanılarak gerçekleştirilmiştir. Mesafe ölçüsü olarak, veri setinin normal dağılım göstermemesi nedeniyle "Manhattan" mesafesi (mutlak farkların toplamı) kullanılmıştır (Aggarwal vd., 2001). Anlamlılık düzeyi $\alpha = 0,05$ olarak kabul edilmiştir.

PERMANOVA testinde, bağımsız değişkenin bağımlı değişken üzerindeki varyansın ne kadarını açıkladığını gösteren R^2 değeri de hesaplanmıştır. R^2, bir bağımsız değişkenin varyansın yüzdesel olarak ne kadarını açıkladığını gösteren bir istatistiktir ve 0 ile 1 arasında bir değer alır. R^2 değeri 1'e yaklaştıkça, bağımsız değişkenin bağımlı değişken üzerindeki açıklayıcı gücü artmaktadır. (Cohen, 1988).

Ek 9’da verilen sonuçlara göre PERMANOVA sonuçları değerlendirilmiş ve değerlendirmeler aşağıda paylaşılmıştır.

WNDCG açısından etkiler Ek 9’da verilen sonuçlara göre değerlendirildiğinde aşağıdaki sonuçlara ulaşılmıştır.

Komut tipinin WNDCG üzerinde istatistiksel olarak anlamlı bir etkisi bulunmamaktadır ($p = 0,749$, $R^2 = 0,002$, Ek 9). Bu, zincirleme düşünce, basit komut vb. farklı komut tiplerinin öneri sisteminin sıralama kalitesi üzerinde belirgin bir fark yaratmadığını

göstermektedir. Komut tipi değişiklikleri, WNDCG varyansının sadece %0,2'sini açıklamaktadır.

GPT sürümünün WNDCG üzerinde istatistiksel olarak anlamlı bir etkisi bulunmaktadır ($p < 0,001$, $R^2 = 0,035$, Ek 9). Bu sonuç, daha gelişmiş bir GPT sürümü olan GPT-4o'nun kullanılmasıyla sıralama kalitesinin artabileceğini göstermektedir. GPT sürümündeki değişiklikler, WNDCG varyansının %3,5'ini açıklamaktadır.

Malzeme seçiminin WNDCG üzerinde istatistiksel olarak anlamlı bir etkisi bulunmamaktadır ($p = 0,063$, $R^2 = 0,003$, Ek 9). Kullanılan malzeme seçim yönteminin (örneğin, belirli mutfak türüne ait veya mevsimsel malzemeler) sıralama kalitesi üzerinde belirgin bir fark yaratmadığı anlaşılmaktadır. Bu değişiklikler, WNDCG varyansının yalnızca %0,3'ünü açıklamaktadır.

GPT sıcaklık değerinin WNDCG üzerinde istatistiksel olarak anlamlı bir etkisi bulunmamaktadır ($p = 0,258$, $R^2 < 0,001$, Ek 9). Bu, modelin yaratıcılığını kontrol eden sıcaklık parametresinin sıralama kalitesi üzerinde belirgin bir rol oynamadığını göstermektedir.

En beğenilen N adet tarifin WNDCG üzerinde istatistiksel olarak anlamlı bir etkisi bulunmaktadır ($p = 0,01$, $R^2 = 0,004$, Ek 9). Bu, kullanıcı geri bildirimlerinin (beğenilen tarifler) sıralama kalitesini etkileyebileceğini ortaya koysa da etkisi toplam varyansın yalnızca %0,4'ünü açıklayacak kadar küçüktür.

LLM tarafından önerilmesi beklenen tarif adedinin WNDCG üzerinde istatistiksel olarak anlamlı ve en büyük etkiye sahip olduğu tespit edilmiştir ($p < 0,001$, $R^2 = 0,214$, Ek 9). Bu, modele girdi olarak verilen LLM tarafından önerilmesi beklenen tarif adedinin sıralama kalitesi üzerinde güçlü bir etki yarattığını göstermektedir. LLM tarafından önerilen tarif adedindeki değişiklikler, toplam varyansın %21,4'ünü açıklamaktadır.
NDCG açısından Ek 9'da verilen sonuçlara göre yapılan değerlendirmelere göre sonuçlar aşağıdaki gibidir.

Komut tipinin NDCG üzerinde istatistiksel olarak anlamlı bir etkisi bulunmamaktadır ($p = 0,814$, $R^2 = 0,002$, Ek 9). Bu, farklı komut türlerinin öneri sisteminin sıralama kalitesinde belirgin bir fark yaratmadığını, değişikliklerin NDCG varyansının sadece %0,2'sini açıkladığını göstermektedir.

GPT sürümünün NDCG üzerinde ise istatistiksel olarak anlamlı bir etkisi olmadığı gözlemlenmiştir ($p = 0,978$, $R^2 < 0,001$, Ek 9), yani kullanılan GPT sürümü sıralama kalitesinde belirgin bir fark oluşturamamaktadır.

Malzeme seçimi ise NDCG üzerinde istatistiksel olarak anlamlı etkiye sahiptir ($p < 0,001$, $R^2 = 0,070$, Ek 9). Bu durum, kullanılan malzeme seçim yönteminin sıralama kalitesini etkilediğini ve varyansın %7'sini açıkladığını ortaya koymaktadır.

GPT sıcaklık değeri de NDCG üzerinde istatistiksel olarak anlamlı bir etki göstermemektedir ($p = 0,753$, $R^2 < 0,001$, Ek 9).

En beğenilen N adet tarifin NDCG üzerinde istatistiksel olarak anlamlı bir etkisi olmadığı belirlenmiştir ($p = 0,330$, $R^2 < 0,001$, Ek 9).

LLM tarafından önerilmesi beklenen tarif adedi, NDCG üzerinde en büyük etkiye sahiptir ($p < 0,001$, $R^2 = 0,456$, Ek 9). Modele girilen tarif sayısı, sıralama kalitesini güçlü bir şekilde etkilemektedir. LLM tarafından önerilen tarif adedindeki değişiklikler, NDCG varyansının %45,6'sını açıklamaktadır.

RMSE açısından etkiler Ek 9'da verilen sonuçlara göre değerlendirilmiş ve sonuçlar aşağıda paylaşılmıştır.

Komut tipinin RMSE üzerinde istatistiksel olarak anlamlı etkisi bulunmamaktadır ($p = 0,450$, $R^2 = 0,003$, Ek 9); yani farklı komut tiplerinin tahmin hatası üzerinde belirgin bir fark yaratmadığı anlaşılmaktadır.

Buna karşın, GPT sürümünün RMSE üzerinde istatistiksel olarak anlamlı bir etkisi vardır ($p < 0,001$, $R^2 = 0,020$, Ek 9). Daha gelişmiş bir GPT sürümünün kullanılması, RMSE'yi düşürerek tahmin doğruluğunu artırabilmekte ve bu değişiklik, RMSE varyansının %2'sini açıklamaktadır.

Ayrıca, malzeme seçimi yöntemi de RMSE üzerinde istatistiksel olarak anlamlı etki göstermektedir ($p < 0,001$, $R^2 = 0,036$, Ek 9); malzeme seçimi değişiklikleri, RMSE varyansının %3,6'sını açıklayarak tahmin doğruluğunu etkilemektedir.

GPT sıcaklık değerinin ise RMSE üzerinde istatistiksel olarak anlamlı bir etkisi bulunmamaktadır ($p = 0,587$, $R^2 < 0,001$, Ek 9).

En beğenilen N adet tarifin RMSE üzerinde istatistiksel olarak anlamlı etkisi vardır ($p = 0,008$, $R^2 = 0,005$, Ek 9).
Son olarak, LLM tarafından önerilmesi beklenen tarif adedi RMSE üzerinde en büyük ve istatistiksel olarak anlamlı etkiye sahiptir ($p < 0,001$, $R^2 = 0,381$, Ek 9). Modele girilen tarif sayısındaki artış, RMSE'yi düşürerek modelin tahmin doğruluğunu önemli ölçüde iyileştirmekte ve toplam varyansın %38,1'ini açıklamaktadır.

MAE açısından etkiler Ek 9'da verilen sonuçlara göre değerlendirilmiş ve sonuçlar aşağıda verilmiştir.

Komut tipinin MAE üzerinde istatistiksel olarak anlamlı bir etkisi bulunmamaktadır ($p = 0,414$, $R^2 = 0,004$, Ek 9); yani farklı komut tiplerinin sistemin ortalama mutlak hatası üzerinde belirgin bir fark yaratmadığı gözlemlenmektedir.

Buna karşın, GPT sürümünün MAE üzerinde istatistiksel olarak anlamlı etkisi vardır ($p < 0,001$, $R^2 = 0,033$, Ek 9). Bu durum, daha gelişmiş bir GPT sürümünün kullanılmasıyla hata oranının düşürülebileceğini göstermektedir.

Malzeme seçimi yöntemi de MAE üzerinde istatistiksel olarak anlamlı etki göstermektedir ($p < 0,001$, $R^2 = 0,027$, Ek 9); yani seçilen malzeme yöntemi sistemin hata oranını etkilemektedir.

Öte yandan, GPT sıcaklık değerinin MAE üzerinde istatistiksel olarak anlamlı bir etkisi bulunmamaktadır ($p = 0,492$, $R^2 < 0,001$, Ek 9).

En beğenilen N adet tarifin MAE üzerinde istatistiksel olarak anlamlı bir etkisi saptanmıştır ($p = 0,012$, $R^2 < 0,001$, Ek 9).

En belirleyici unsur ise LLM tarafından önerilmesi beklenen tarif adedidir ($p < 0,001$, $R^2 = 0,202$, Ek 9). Daha fazla tarif içeren bir veri kümesi, modelin MAE performansını önemli ölçüde iyileştirerek hata oranını düşürmektedir.

Yanıt süresi açısından etkiler değerlendirildiğinde aşağıdaki sonuçlara ulaşılmıştır.

Komut tipinin yanıt süresi üzerinde istatistiksel olarak anlamlı bir etkisi bulunmamaktadır ($p < 0,312$, $R^2 = 0,004$, Ek 9).

GPT sürümünün yanıt süresi üzerinde istatistiksel olarak anlamlı bir etkisi bulunmaktadır ($p < 0,001$, $R^2 = 0,165$, Ek 9). Daha gelişmiş GPT sürümleri, daha karmaşık işlemler gerektirdiğinden yanıt süresini etkileyebilmektedir ve model mimarisinin yanıt süresi üzerinde önemli rol oynadığını göstermektedir.

Malzeme seçimi ise yanıt süresi üzerinde istatistiksel olarak anlamlı etki göstermemektedir ($p = 0,435$, $R^2 < 0,001$, Ek 9).

GPT sıcaklık değeri de yanıt süresi üzerinde istatistiksel olarak anlamlı etki göstermektedir ($p = 0,046$, $R^2 = 0,003$, Ek 9). Bu değer, modelin çıktı üretme sürecini etkileyerek yanıt süresinde değişikliklere yol açabilmektedir.

En beğenilen N adet tarifin yanıt süresi üzerinde istatistiksel olarak anlamlı etkisi olmadığı belirlenmiştir ($p = 0,490$, $R^2 < 0,001$, Ek 9).

Son olarak, LLM tarafından önerilen tarif adedinin yanıt süresi üzerinde istatistiksel olarak anlamlı bir etkisi vardır ($p < 0,001$, $R^2 = 0,028$, Ek 9). Daha fazla tarif içeren veri kümeleri, modelin daha fazla veriyi işlemesi gerektiğinden daha uzun yanıt sürelerine neden olmaktadır.

Toplam jeton adedi açısından yapılan değerlendirmelere göre sonuçlar aşağıdaki gibidir.

Toplam jeton adedi, öneri sisteminin kullandığı kaynak miktarını gösterir; daha az jeton kullanımı genellikle daha düşük maliyet ve daha yüksek verimlilik anlamına gelmektedir.

Komut tipi, toplam jeton adedi üzerinde istatistiksel olarak anlamlı etkiye sahiptir ($p < 0,001$, $R^2 = 0,111$, Ek 9). Farklı komut tipleri, modelin işlediği kelime ve karakter miktarını değiştirerek jeton tüketimini etkileyebilir. Özellikle zincirleme düşünce içeren gibi daha karmaşık komutlar, daha fazla jeton tüketimine yol açar ve bu değişiklikler toplam varyansın %11,1'ini açıklamaktadır.

GPT sürümü de toplam jeton adedi üzerinde istatistiksel olarak anlamlı etkiye sahiptir ($p < 0,001$, $R^2 = 0,083$, Ek 9). Daha gelişmiş GPT sürümleri, daha karmaşık mimarileri ve daha fazla parametreleri nedeniyle daha fazla işlem gücü gerektirir; bu da daha fazla jeton kullanımına neden olur. GPT sürümü, toplam jeton adedi varyansının %8,3'ülük bir dilimini açıklayabilmektedir.

Malzeme seçimi yöntemi ise toplam jeton adedi üzerinde istatistiksel olarak anlamlı etki gösterse de ($p = 0,012$, $R^2 = 0,005$, Ek 9) varyansın sadece %0,5'ini bu faktör ile açıklanmaktadır.

GPT sıcaklık değeri, toplam jeton adedi üzerinde istatistiksel olarak anlamlı bir etkiye sahiptir ($p = 0,031$, $R^2 = 0,004$, Ek 9). Daha yüksek sıcaklık değerleri, modelin daha çeşitli

ve potansiyel olarak daha uzun çıktılar üretmesine yol açarak jeton tüketimini artırabilir; ancak bu etkinin varyans payı %0,4 ile oldukça düşüktür.

En beğenilen N adet tarifin toplam jeton kullanımı üzerinde istatistiksel olarak anlamlı bir etkisi bulunmamaktadır ($p = 0,746$, $R^2 < 0,001$, Ek 9).

En büyük etki ise, modele girdi olarak LLM tarafından önerilmesi beklenen tarif adedinde görülmektedir ($p < 0,001$, $R^2 = 0,270$, Ek 9). Tarif adedindeki artış, işlenecek veri miktarını artırarak jeton tüketimini yükseltmekte ve toplam varyansın %27'sini açıklamaktadır.

Bu sonuçlara göre, ÜYZ tabanlı yemek tarifi öneri sistemlerinin geliştirilmesinde öncelikle LLM tarafından önerilmesi beklenen tarif adedine odaklanılması gerekmektedir. Daha fazla ve farklı türde tarif içeren sonuçların LLM'den talep edilmesi, model performansını önemli ölçüde etkilemektedir.

Ayrıca, GPT sürümü, malzeme seçimi, komut tipi ve GPT sıcaklık değeri gibi diğer faktörlerin optimizasyonu, özellikle yanıt süresi veya jeton tüketimi gibi belirli metriklerin iyileştirilmesi hedefleniyorsa fayda sağlayabilir.

Gelecek çalışmalarda, bu faktörlerin etkileşimleri detaylı şekilde incelenmeli, farklı ÜYZ mimarileri karşılaştırılmalı ve kullanıcı geri bildirimleri modelleme sürecine dahil edilmelidir; böylece ÜYZ'nin öneri sistemlerindeki potansiyeli tam olarak ortaya konulabilir.

Son olarak, bazı bağımsız değişkenler için düşük R^2 değerleri, modelin açıklayamadığı varyansın doğrusal olmayan ilişkilerden, model dışı faktörlerden veya veri setindeki doğal değişkenlikten kaynaklanabileceğini göstermektedir.

4.6. Tartışma

ÜYZ tabanlı bir yemek tarifi öneri sisteminin performansı, Kruskal-Wallis, Epsilon Kare ve PERMANOVA analizleri ile incelemiştir.

Elde edilen bulgular, öneri sistemlerinin karmaşık sistemler olduğunu ve başarılarının, tek bir faktörün optimizasyonundan ziyade, birden fazla faktörün etkileşimine bağlı olduğunu göstermektedir. Bu çok faktörlü yapı, ÜYZ'nin "kara kutu" (black box) doğası ve farklı bileşenlerin (model mimarisi, eğitim verisi, girdi parametreleri) bir araya gelerek performansı şekillendirmesi ile tutarlı gözükmektedir.

Analizler, veri miktarının (LLM tarafından önerilmesi beklenen tarif adedi) kritik önemini vurgulamaktadır. Tüm istatistiksel yöntemlerde, LLM tarafından önerilmesi beklenen tarif adedinin, sıralama kalitesi (WNDCG, NDCG) ve tahmin doğruluğu (RMSE, MAE) gibi temel performans metrikleri üzerinde en güçlü ve tutarlı etkiye sahip olduğu gözlemlenmiştir. Bu durum, ÜYZ modellerinin "veri oburluğu" (data hungriness) özelliğiyle (Brown vd., 2020) örtüşmektedir.

Testlerde incelenen diğer faktörler (GPT sürümü, malzeme seçimi, komut tipi, GPT sıcaklık değeri, kullanıcının beğendiği LLM tarafından önerilmesi beklenen tarif adedi) ise, sistem performansının farklı boyutları üzerinde, değişen derecelerde etkilere sahiptir. Bu faktörlerin etkileri, hem kullanılan istatistiksel yönteme (Kruskal-Wallis, Epsilon Kare, PERMANOVA) hem de değerlendirilen performans metriğine (WNDCG, NDCG, RMSE, MAE, Yanıt Süresi, Toplam Jeton Adedi) göre farklılık göstermiştir.

Örneğin, daha gelişmiş GPT sürümleri, genellikle sıralama kalitesini ve tahmin doğruluğunu artırırken, yanıt süresini ve jeton tüketimini de etkileyebilmektedir. Malzeme seçimi, sıralama kalitesini ve tahmin hatasını; komut tipi, hata oranını, yanıt süresini ve jeton tüketimini; GPT sıcaklık değeri ise yanıt süresini ve jeton tüketimini etkilemektedir. Bu bulgular, öneri sistemlerinin çok amaçlı optimizasyon gerektiren sistemler olduğunu göstermektedir. Tek bir metriği (örneğin, sadece sıralama kalitesini)

optimize etmeye çalışmak, diğer metriklerde (örneğin, yanıt süresi veya maliyet) istenmeyen sonuçlara yol açabilmektedir.

Sonuç olarak, ÜYZ tabanlı öneri sistemlerinin performansının, LLM tarafından önerilmesi beklenen tarif adedi, model seçimi ve girdi parametrelerinin dikkatli bir şekilde ayarlanmasıyla optimize edilebileceği gösterilmiştir. Özellikle, veri setinin büyüklüğü ve çeşitliliği, sistemin başarısı için kritik öneme sahiptir. Gelecek çalışmalarda, bu faktörlerin etkileşimlerini (interaction effects) daha derinlemesine inceleyen (örneğin, farklı GPT sürümlerinin farklı tarif adetlerinde nasıl performans gösterdiğini araştıran) ve kullanıcı geri bildirimlerini de modelleme sürecine daha etkin bir şekilde dahil eden çalışmalar, ÜYZ'nin öneri sistemlerindeki potansiyelini daha da açığa çıkaracaktır.

Bununla birlikte, özellikle PERMANOVA analizinde karşılaşılan düşük R^2 değerleri, önemli bir noktaya işaret etmektedir, önerilen modelin açıklayamadığı bir varyans payı mevcuttur. Bu durum, ÜYZ performansını etkileyen faktörlerin, basit doğrusal ilişkilerden ibaret olmadığını ve daha karmaşık dinamiklerin (doğrusal olmayan etkileşimler, gizli değişkenler, anlık sistem yoğunluğu vb.) rol oynadığını göstermektedir.

Gelecekte, bu karmaşıklığı daha iyi yakalayabilmek adına, doğrusal olmayan modellere, daha zengin veri setlerine, öznitelik (feature) mühendisliğine ve farklı istatistiksel analizlere yönelmek, modelin açıklama gücünü ve dolayısıyla öneri sisteminin başarısını artırmada kritik öneme sahip olacaktır.

5. SONUÇ

Bu tezde, büyük dil modellerinin (LLM) yemek tarifi öneri sistemlerine entegrasyonundaki potansiyel faydaları ve kısıtları, çok yönlü bir yaklaşımla incelenmiştir. Çalışmanın temelinde, LLM'lerin, geleneksel öneri yöntemlerinin ötesine geçerek, bağlamsal anlama, kişiselleştirme ve yaratıcılık yetenekleriyle yemek tarifi önerilerini nasıl iyileştirebileceği sorusu yatmaktaydı. Bu bağlamda, farklı komut stratejilerinin, LLM sürümlerinin ve malzeme seçim yöntemlerinin öneri performansı üzerindeki etkileri analiz edilmiş; aynı zamanda üretken yapay zekânın bu alandaki sınırlamaları ve gelecek araştırma yönleri değerlendirilmiştir.

Deneysel analizler ve istatistiksel testler (Kruskal-Wallis ve PERMANOVA), tezin ana hedeflerine ulaşıldığını göstermektedir. LLM'lerin yemek tarifi öneri sistemlerinde kullanımı, özellikle girdi olarak LLM'nin önerip değerlendireceği tarif sayısının yüksek olması durumunda, öneri performansını (sıralama kalitesi ve tahmin doğruluğu) önemli ölçüde artırmıştır. Bu durum, LLM'lerin kullanıcı tercihlerini ve tarif özelliklerini daha iyi anlayarak, daha isabetli ve kişiselleştirilmiş öneriler sunabildiğini göstermektedir. Farklı komut tipleri ile GPT sürümlerinin sistem performansı üzerindeki etkileri de detaylı bir şekilde incelenmiş; her bir stratejinin farklı performans metrikleri üzerinde farklı etkilere sahip olduğu görülerek, sistem optimizasyonu için bu faktörlerin değerlendirilmesi gerektiği ortaya konmuştur.

Bu tez, özellikle LLM'nin önerip değerlendireceği tarif sayısının (LLM tarafından önerilmesi beklenen tarif adedi), yemek tarifi öneri sisteminin başarısı üzerinde en kritik faktör olduğunu vurgulamaktadır. Bu bulgu, LLM tabanlı sistemlerin, kullanıcıların damak zevklerini ve beslenme tercihlerini daha kapsamlı şekilde ele alarak daha alakalı ve kişiselleştirilmiş tarifler sunabileceğini göstermektedir. Çalışma aynı zamanda, LLM tabanlı öneri sistemlerinin mevcut sınırlamalarını ("kara kutu" doğası, yüksek hesaplama maliyetleri, yanıt süresi dalgalanmaları ve açıklanamayan varyans) ortaya koymuş ve bu sınırlamaların aşılması, sistemlerin daha da geliştirilmesi için gelecekte yapılması gereken araştırmaları detaylı bir şekilde tartışmıştır.

Sonuç olarak, bu çalışma ÜYZ ve LLM'lerin yemek tarifi öneri sistemlerinde kullanımının, geleneksel yöntemlere kıyasla önemli avantajlar sunduğunu, ancak bu teknolojinin potansiyelinden tam olarak yararlanmak için dikkatli bir optimizasyon ve daha fazla araştırma gerektiğini ortaya koymuştur. Elde edilen bulgular hem akademik araştırmacılara hem de sektör profesyonellerine, daha iyi yemek tarifi öneri sistemleri tasarlamaları için değerli bilgiler ve yol gösterici ilkeler sunmaktadır.

Bu tez, LLM tabanlı yemek tarifi öneri sistemlerinin geleceğini şekillendirecek yeni araştırmalara zemin hazırlamış; özellikle daha fazla veri, gelişmiş modeller ve kullanıcı odaklı yaklaşımlarla bu sistemlerin daha da geliştirilebileceğini göstermiştir.

KAYNAKLAR

Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). *On the surprising behavior of distance metrics in high dimensional space*. In International conference on database theory (pp. 420-434). Springer, Berlin, Heidelberg.

Anderson, M. J. (2001). *A new method for non-parametric multivariate analysis of variance*. Austral Ecology, 26(1), 32-46.

Anderson, T. W. & Darling, D. A. (1954). *A test of goodness of fit*. Journal of the American Statistical Association, 49(268), 765–769. https://doi.org/10.1080/01621459.1954.10501232

Bartlett, M. S. (1937). *Properties of sufficiency and statistical tests*. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 160(901), 268–282. https://doi.org/10.1098/rspa.1937.0109

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P. & Amodei, D. (2020). *Language models are few-shot learners*. Advances in Neural Information Processing Systems, 33, 1877–1901.

Büyüköztürk, Ş., Çokluk, Ö., & Köklü, N. (2018). *SPSS ile biyoistatistik: Parametrik olmayan yöntemler*. Pegem Akademi Yayınları.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2. ed.).* Academic Press.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K. & Xie, X. (2023). *A survey on evaluation of large language models*. arXiv. https://arxiv.org/abs/2307.03109

Dai, S., Shao, N., Zhao, H., Yu, W., Si, Z., Xu, C., Sun, Z., Zhang, X. & Xu, J. (2023). *Uncovering ChatGPT's capabilities in recommender systems*. In Proceedings of RecSys (ss. 1126–1132). ACM.

Di Palma, D., Biancofiore, G. M., Anelli, V. W., Narducci, F., Di Noia, T., & Di Sciascio, E. (2023). *Evaluating ChatGPT as a recommender system: A rigorous approach [Preprint].* arXiv. https://doi.org/10.48550/arXiv.2309.03613

Fan, W., Zhao, Z., Li, J., Liu, Y., Mei, X., Wang, Y., Tang, J. & Li, Q. (2023). *Recommender systems in the era of large language models (LLMs)*. arXiv. https://arxiv.org/abs/2307.02046

Field, A. (2018). *Discovering statistics using IBM SPSS statistics (5. ed.)*. Sage.

Friedman, L., Ahuja, S., Allen, D., Tan, T., Sidahmed, H., Long, C., Xie, J., Schubiner, G., Patel, A. & Lara, H. (2023). *Leveraging large language models in conversational recommender systems*. arXiv. https://arxiv.org/abs/2305.07961

Harte, J., Zorgdrager, W., Louridas, P., Katsifodimos, P., Jannach, P. & Fragkoulis, M. (2023). *Leveraging large language models for sequential recommendation*. In Proceedings of RecSys (ss. 1096–1102). ACM.

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction (2. ed.)*. Springer.

Hollander, M., Wolfe, D. A., & Chicken, E. (2013). *Nonparametric statistical methods*. John Wiley & Sons.

Hu, L. (2023). *Generative AI and future. Towards AI*. https://pub.towardsai.net/generativeai-and-future-c3b1695876f2

Jhamat, N., Mustafa, G., Arshad, Z. & Abbas, R. (2021). *Expert system for recommendations of healthy food recipes using machine learning*. İlköğretim Online–Elementary Education Online, 20(5), 2867–2874.

Jovanović, M. (2023). *Generative artificial intelligence: Trends and prospects*. Computer, 10(99), 1–10. https://www.computer.org/csdl/magazine/co/2022/10/09903869/1H0G6xvtREk.0.1109/MC.2022.3192720

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T. & Kasneci, G. (2023). *ChatGPT for good? On opportunities and challenges of large language models for education*. Learning and Individual Differences, 103.

Kruskal, W. H., & Wallis, W. A. (1952). *Use of ranks in one-criterion variance analysis.* Journal of the American Statistical Association, 47(260), 583–621. 1 https://doi.org/10.1080/01621459.1952.10483441

Kumar, C., Chowdary, C. R. & Meena, A. K. (2024). *Recent trends in recommender systems: A survey*. International Journal of Multimedia Information Retrieval, 13(1), 41. https://doi.org/10.1007/s13735-024-00349-1

Lanling, X., Junjie, Z., Bingqian, L., Jinpeng, W., Mingchen, C., Wayne, X. Z. & Ji-Rong, W. (2023). *Prompting large language models for recommender systems: A comprehensive framework and empirical analysis*. In Proceedings of ACM.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, F. Mosteller, & J. W. Tukey (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 278–292). Stanford University Press.

Li, S. & Majumder, B. P. (2020). *Food.com recipes and interactions*. Kaggle. https://www.kaggle.com/datasets/shuyangli94/food-com-recipes-and-user-interactions

Li, Y., Liu, K., Satapathy, R., Wang, S. & Cambria, E. (2024). *Recent developments in recommender systems: A survey*. IEEE Computational Intelligence Magazine, 19(2), 78–95. https://doi.org/10.1109/MCI.2024.3363984

Liu, J., Liu, C., Zhou, P., Lv, R., Zhou, K., & Zhang, Y. (2023). Is ChatGPT a good recommender? A preliminary study. CoRR, abs/2304.10149. https://doi.org/10.48550/arXiv.2304.10149

Lohr, S. (2009). *Netflix awards $1 million prize and starts a new contest*. The New York Times. https://www.nytimes.com/2009/09/22/technology/internet/22netflix.html

McArdle, B. H., & Anderson, M. J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology, 82(1), 290-297.

Montgomery, D. C. (2017). *Design and analysis of experiments (9th ed.)*. Wiley.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., & Wagner, H. (2020). *vegan: Community Ecology Package*. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

Özdamar, K. (2019). *Paket programlar ile istatistiksel veri analizi*. Kaan Kitabevi.

Peng, L., Lemei, Z. & Atle, J. G. (2023). *Pre-train, prompt, and recommendation: A comprehensive survey of language modeling paradigm adaptations in recommender systems*. Transactions of the Association for Computational Linguistics, 11, 1553–1571.

Ricci, F., Rokach, L. & Shapira, B. (2010). *Recommender systems handbook*. Springer. https://doi.org/10.1007/978-0-387-85820-3

Shapiro, S. S. & Wilk, M. B. (1965). *An analysis of variance test for normality (complete samples).* Biometrika, 52(3/4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences (2nd ed.)*. McGraw-Hill.

Stephens, M. A. (1974). *EDF statistics for goodness of fit and some comparisons*. Journal of the American Statistical Association, 69(347), 730–737.

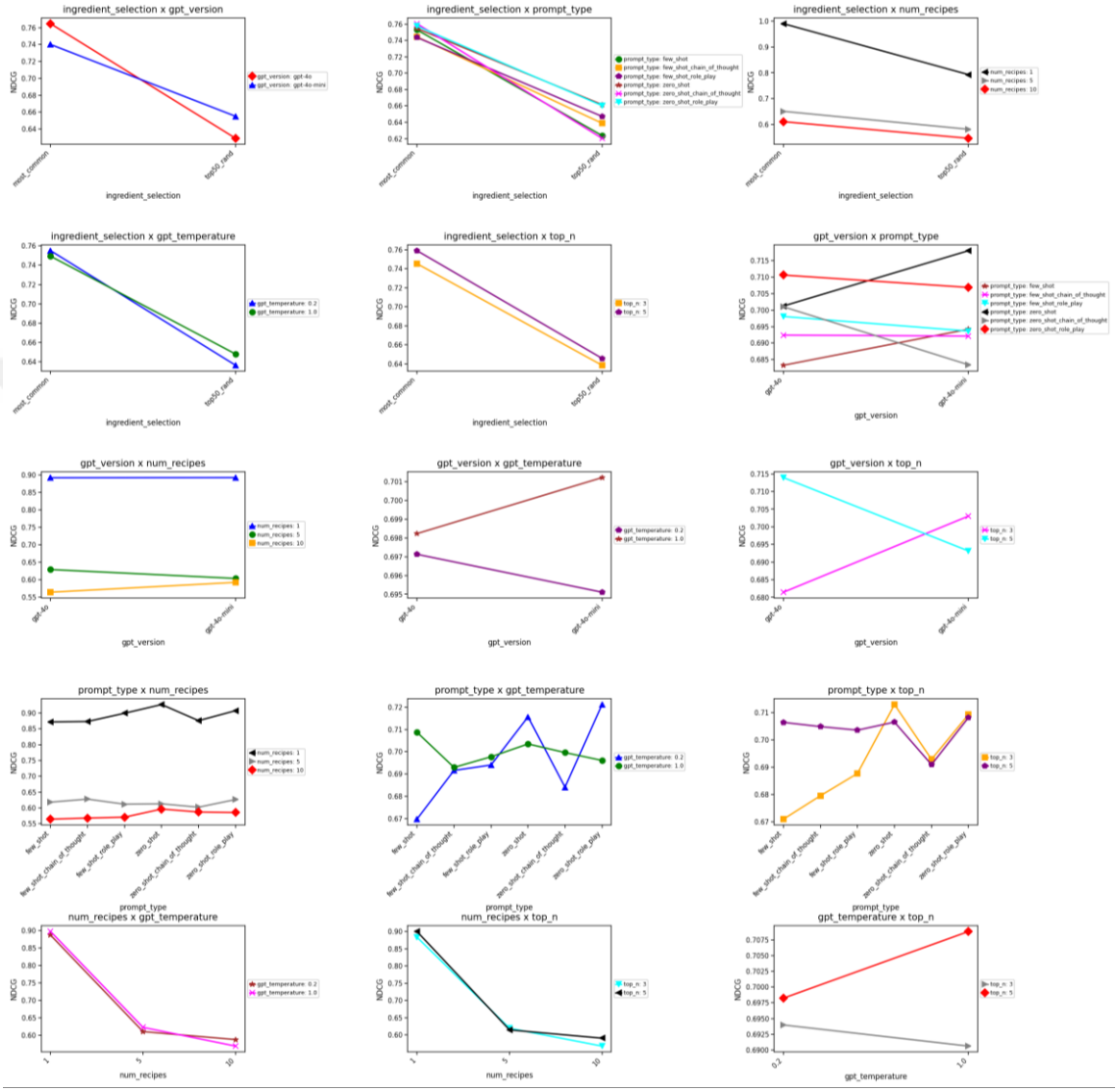
Tamm, Y. M., Damdinov, R., & Vasilev, A. (2021). *Quality metrics in recommender systems: Do we calculate metrics consistently?* In Proceedings of the 15th ACM conference on recommender systems (pp. 708-713).

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J. & Schmidt, D. C. (2023). *A prompt pattern catalog to enhance prompt engineering with ChatGPT*. arXiv. https://arxiv.org/abs/2302.11382

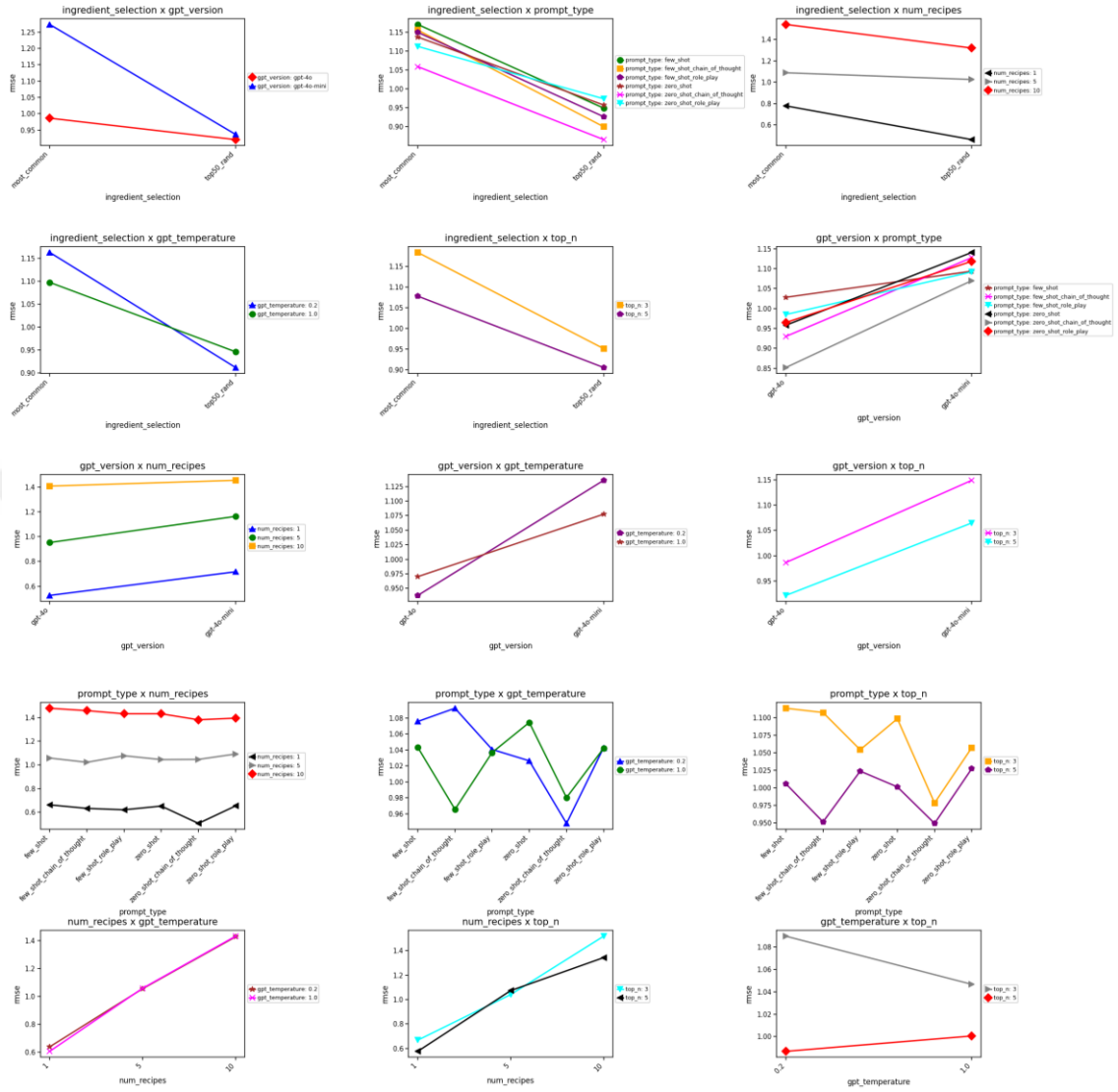
Xu, L., Zhang, J., Li, B., Wang, J., Cai, M., Zhao, W. X. & Wen, J.-R. (2018). *Prompting large language models for recommender systems: A comprehensive framework and empirical analysis*. In ACM.

EKLER

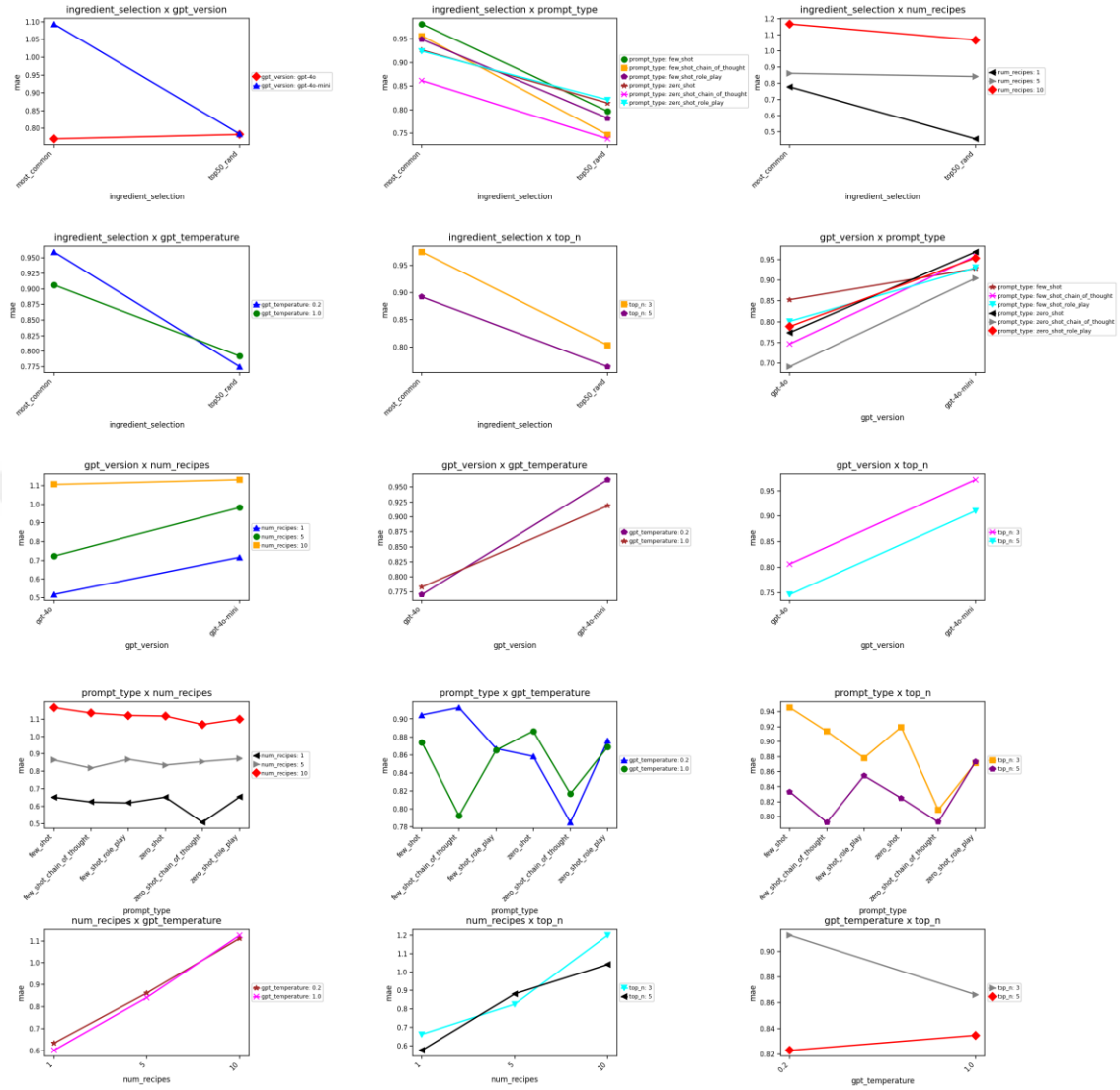
EK 1 NDCG etkileşim grafikleri



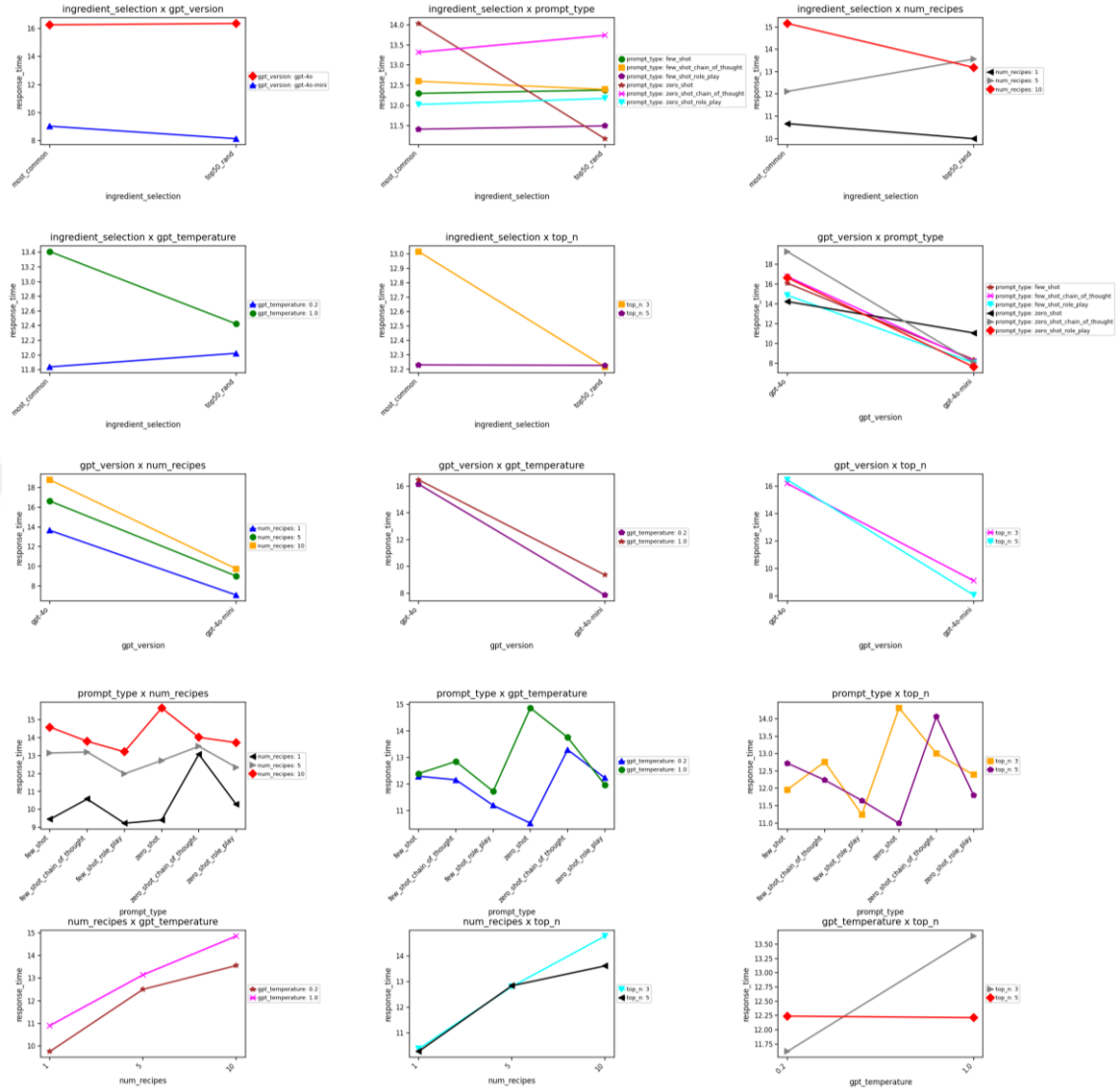
EK 2 RMSE etkileşim grafikleri



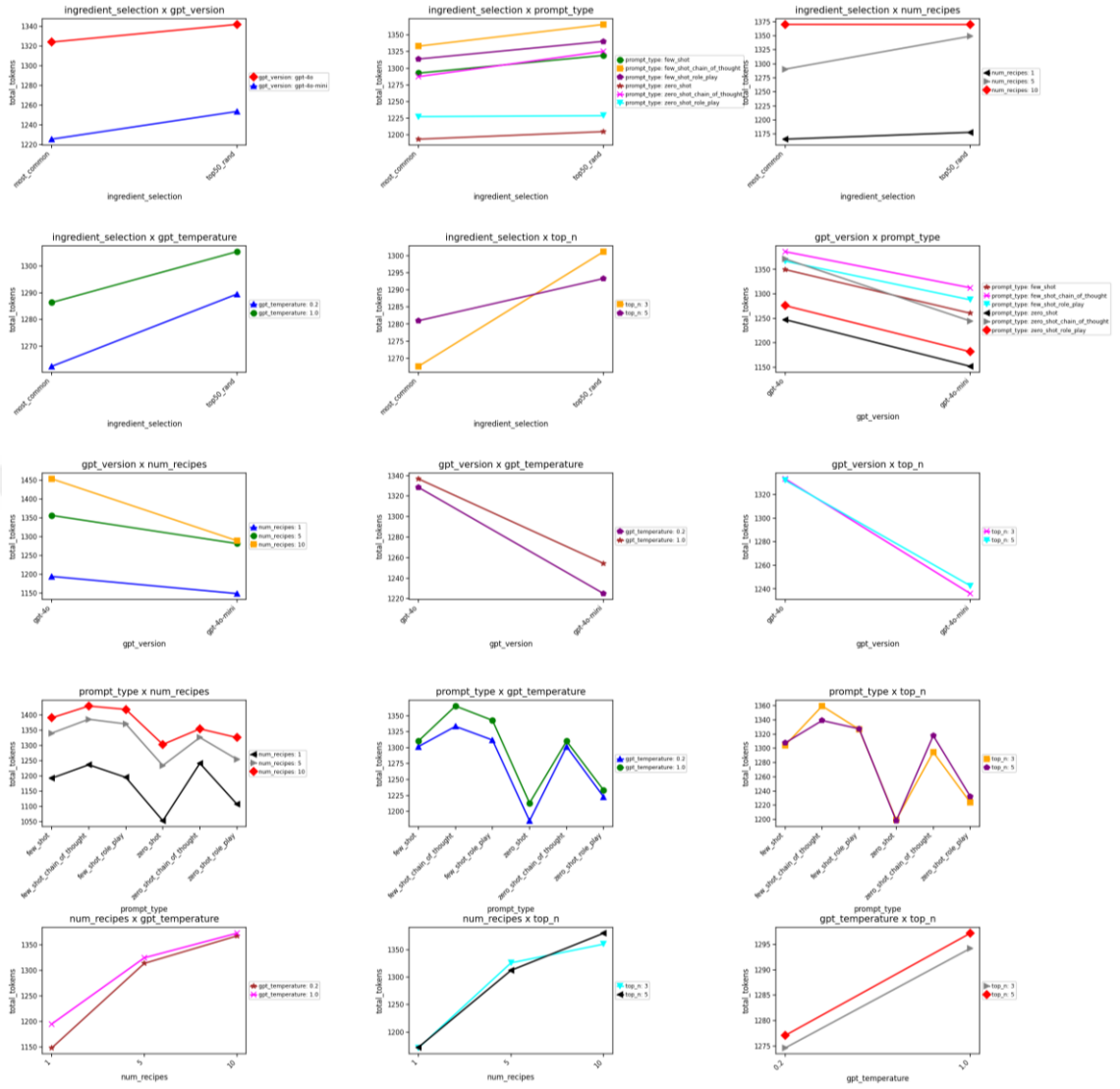
EK 3 MAE etkileşim grafikleri



EK 4 Yanıt süresi etkileşim grafikleri



EK 5 Toplam jeton adedi etkileşim grafikleri



EK 6 Shapiro–Wilk testi sonuçları (p-degeri ≤ 0,05 ise koyu olarak işaretlenmiştir)

Değişken	Shapiro–Wilk p-değeri	Değerlendirme
WNDCG	**<0,001***	Normallik varsayımı reddedildi.
NDCG	**<0,001***	Normallik varsayımı reddedildi.
RMSE	**<0,001***	Normallik varsayımı reddedildi.
MAE	**<0,001***	Normallik varsayımı reddedildi.
Yanıt Süresi	**<0,001***	Normallik varsayımı reddedildi.
Toplam Jeton Adedi	**<0,001***	Normallik varsayımı reddedildi.

Shapiro–Wilk istatistiği için elde edilen p-değeri, seçilen anlamlılık düzeyi ($\alpha = 0,05$) ile karşılaştırılmıştır. Eğer p-değeri 0,05'ten küçükse, normal dağılım varsayımının reddedildiği; yani verilerin normal dağılıma uymadığı sonucuna varılmaktadır. (Shapiro & Wilk, 1965). Bu test için α (anlamlılık düzeyi), 0,05 olarak belirlenmiştir.

EK 7 Anderson–Darling testi sonuçları ($\boldsymbol{A^2} \geq 0{,}787$ ise koyu olarak işaretlenmiştir)

Değişken	Anderson–Darling Değeri ($\boldsymbol{A^2}$)	Değerlendirme
WNDCG	**67,072***	Normallik varsayımı reddedildi.
NDCG	**80,218***	Normallik varsayımı reddedildi.
RMSE	**37,080***	Normallik varsayımı reddedildi.
MAE	**55,044***	Normallik varsayımı reddedildi.
Yanıt Süresi	**85,307***	Normallik varsayımı reddedildi.
Toplam Jeton Adedi	**5,279***	Normallik varsayımı reddedildi.

Anderson–Darling (Anderson ve Darling, 1954) testinde, elde edilen A^2 değeri, belirlenen anlamlılık düzeyine ($\alpha = 0{,}05$) karşılık gelen kritik değer ile karşılaştırılmıştır. Literatürde yaygın olarak, $\alpha = 0{,}05$ için kritik değerin yaklaşık 0,787 civarında olduğu belirtilir (Stephens, 1974). Bu durumda, hesaplanan A^2 değeri bu kritik değerin üzerinde ise, verilerin normal dağılım gösterdiği varsayımı reddedilir; aksi halde kabul edilmektedir. Bu test için anlamlılık sınırı $A^2 = 0{,}787$ olarak belirlenmiştir.

EK 8 Kovaryans matrisi homojenliği test sonuçları (p-degeri ≤ 0,05 ise koyu olarak işaretlenmiştir)

Bağımlı Değişken	Test	Etki	p-değeri	Değerlendirme
WNDCG	Bartlett	Malzeme Seçimi	**<0,001**	Homojen değil.
		GPT Sürümü	**<0,001**	Homojen değil.
		Komut Tipi	0,627	Homojen
		LLM Tarafından Önerilmesi Beklenen Tarif Adedi	**<0,001**	Homojen değil.
		GPT Sıcaklık Değeri	0,654	Homojen
		En Beğenilen N Adet Tarif	**0,022**	Homojen değil.
	Levene	Malzeme Seçimi	**<0,001**	Homojen değil.
		GPT Sürümü	**<0,001**	Homojen değil.
		Komut Tipi	0,877	Homojen
		LLM Tarafından Önerilmesi Beklenen Tarif Adedi	**<0,001**	Homojen değil.
		GPT Sıcaklık Değeri	0,989	Homojen.
		En Beğenilen N Adet Tarif	0,289	Homojen.
NDCG	Bartlett	Malzeme Seçimi	**0,004**	Homojen değil.
		GPT Sürümü	**<0,001**	Homojen değil.
		Komut Tipi	0,984	Homojen.
		LLM Tarafından Önerilmesi Beklenen Tarif Adedi	**<0,001**	Homojen değil.
		GPT Sıcaklık Değeri	0,857	Homojen.
		En Beğenilen N Adet Tarif	0,686	Homojen.
	Levene	Malzeme Seçimi	0,227	Homojen.
		GPT Sürümü	**<0,001***	Homojen değil.
		Komut Tipi	0,778	Homojen.

EK 8 Kovaryans matrisi homojenliği test sonuçları (devam)

Bağımlı Değişken	Test	Etki	p-değeri	Değerlendirme
WNDCG	Bartlett	Malzeme Seçimi	**<0,001**	Homojen değil.
		GPT Sıcaklık Değeri	0,797	Homojen.
		En Beğenilen N Adet Tarif	0,938	Homojen.
RMSE	Bartlett	Malzeme Seçimi	**0,026***	Homojen değil.
		GPT Sürümü	**0,001***	Homojen değil.
		Komut Tipi	0,227*	Homojen.
		LLM Tarafından Önerilmesi Beklenen Tarif Adedi	**<0,001***	Homojen değil.
		GPT Sıcaklık Değeri	0,605	Homojen.
		En Beğenilen N Adet Tarif	**0,049***	Homojen değil.
	Levene	Malzeme Seçimi	0,025	Homojen.
		GPT Sürümü	**<0,001***	Homojen değil.
		Komut Tipi	0,866	Homojen.
		LLM Tarafından Önerilmesi Beklenen Tarif Adedi	**<0,001***	Homojen değil.
		GPT Sıcaklık Değeri	0,927	Homojen.
		En Beğenilen N Adet Tarif	0,356	Homojen.
MAE	Bartlett	Malzeme Seçimi	**<0,001***	Homojen değil.
		GPT Sürümü	**<0,001***	Homojen değil.
		Komut Tipi	**0,046***	Homojen değil.
		LLM Tarafından Önerilmesi Beklenen Tarif Adedi	**<0,001***	Homojen değil.
		GPT Sıcaklık Değeri	0,276	Homojen.
		En Beğenilen N Adet Tarif	0,068	Homojen.
	Levene	Malzeme Seçimi	**<0,001***	Homojen değil.

EK 8 Kovaryans matrisi homojenliği test sonuçları (devam)

Bağımlı Değişken	Test	Etki	p-değeri	Değerlendirme
MAE	Levene	GPT Sürümü	**<0,001***	Homojen değil.
		Komut Tipi	0,703	Homojen.
		LLM Tarafından Önerilmesi Beklenen Tarif Adedi	**<0,001***	Homojen değil.
		GPT Sıcaklık Değeri	0,740	Homojen.
		En Beğenilen N Adet Tarif	0,941	Homojen.
Yanıt Süresi	Bartlett	Malzeme Seçimi	**<0,001***	Homojen değil.
		GPT Sürümü	**<0,001***	Homojen değil.
		Komut Tipi	**<0,001***	Homojen değil.
		LLM Tarafından Önerilmesi Beklenen Tarif Adedi	**<0,001***	Homojen değil.
		GPT Sıcaklık Değeri	**<0,001***	Homojen değil.
		En Beğenilen N Adet Tarif	**<0,001***	Homojen değil.
	Levene	Malzeme Seçimi	0,433	Homojen.
		GPT Sürümü	**<0,001***	Homojen değil.
		Komut Tipi	0,206	Homojen.
		LLM Tarafından Önerilmesi Beklenen Tarif Adedi	**0,008***	Homojen değil.
		GPT Sıcaklık Değeri	0,671	Homojen.
		En Beğenilen N Adet Tarif	0,363	Homojen.
Toplam Jeton Adedi	Bartlett	Malzeme Seçimi	0,241	Homojen.
		GPT Sürümü	**<0,001***	Homojen değil.
		Komut Tipi	0,157	Homojen.
		LLM Tarafından Önerilmesi Beklenen Tarif Adedi	**0,003***	Homojen değil.

EK 8 Kovaryans matrisi homojenliği test sonuçları (devam)

Bağımlı Değişken	Test	Etki	p-değeri	Değerlendirme
Toplam Jeton Adedi	Bartlett	GPT Sıcaklık Değeri	0,275	Homojen.
		En Beğenilen N Adet Tarif	0,245	Homojen.
	Levene	Malzeme Seçimi	0,201	Homojen.
		GPT Sürümü	**<0,001***	Homojen değil.
		Komut Tipi	0,491	Homojen.
		LLM Tarafından Önerilmesi Beklenen Tarif Adedi	**0,027***	Homojen değil.
		GPT Sıcaklık Değeri	0,243	Homojen.
		En Beğenilen N Adet Tarif	0,590	Homojen.

Bartlett testi sonucunda elde edilen p-değeri, seçilen anlamlılık düzeyi ($\alpha = 0,05$) ile karşılaştırılmıştır. Eğer p-değeri 0,05'ten küçükse, gruplar arası varyansların eşit olmadığı sonucuna varılmaktadır. Ancak, Bartlett testi normal dağılım varsayımına duyarlı olduğundan, veriler normal dağılmıyorsa test sonucu yanıltıcı olabilmektedir (Bartlett, 1937).

Levene testi, gruplar arasındaki varyansların eşit olup olmadığını kontrol etmek için kullanılan ve özellikle normal dağılım varsayımının sağlanmadığı durumlarda ANOVA öncesi test olarak tercih edilen bir yöntemdir (Levene, 1960).

Sonuçlar belirlenen anlamlılık düzeyi ($\alpha = 0,05$) çerçevesinde değerlendirilmiştir; eğer p-değeri 0,05'in altında kalıyorsa, gruplar arasındaki varyanslarda istatistiksel olarak anlamlı bir farklılık bulunduğu ve dolayısıyla varyansların eşitliği varsayımının ihlal edildiği sonucuna varılabilmektedir (Levene, 1960).

EK 9 PERMANOVA sonuçları (p-degeri ≤ 0,05 ise koyu olarak işaretlenmiştir)

Bağımlı Değişken	Etki	df	SumOfSqs	R^2	F	p
WNDCG	Komut Tipi	5	0,110	0,002	0,526	0,749
	GPT Sürümü	1	2,032	0,035	50,260	**<0,001***
	Malzeme Seçimi	1	0,147	0,003	3,516	0,063
	GPT Sıcaklık Değeri	1	0,054	<0,001	1,300	0,258
	En Beğenilen N Adet Tarif	1	0,264	0,004	6,322	**0,01***
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	2	12,591	0,214	191,240	**<0,001***
NDCG	Komut Tipi	5	0,097	0,002	0,448	0,814
	GPT Sürümü	1	0,000	<0,001	0,001	0,978
	Malzeme Seçimi	1	4,256	0,070	106,081	**<0,001***
	GPT Sıcaklık Değeri	1	0,005	<0,001	0,107	0,753
	En Beğenilen N Adet Tarif	1	0,044	<0,001	1,012	0,330
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	2	27,615	0,456	588,085	**<0,001***
RMSE	Komut Tipi	5	1,350	0,003	0,942	0,450
	GPT Sürümü	1	8,226	0,020	29,291	**<0,001***
	Malzeme Seçimi	1	14,315	0,036	51,774	**<0,001***
	GPT Sıcaklık Değeri	1	0,077	<0,001	0,267	0,587
	En Beğenilen N Adet Tarif	1	1,966	0,005	6,890	**0,008***
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	2	153,190	0,381	431,383	**<0,001***
MAE	Komut Tipi	5	1,091	0,004	1,055	0,414

EK 9 PERMANOVA sonuçları (devam)

Bağımlı Değişken	Etki	df	SumOfSqs	R^2	F	p
MAE	GPT Sürümü	1	9,484	0,033	47,376	**<0,001***
	Malzeme Seçimi	1	7,866	0,027	39,070	**<0,001***
	GPT Sıcaklık Değeri	1	0,107	<0,001	0,515	0,492
	En Beğenilen N Adet Tarif	1	1,299	<0,001	6,303	**0,012***
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	2	58,708	0,202	177,692	**<0,001***
Yanıt Süresi	Komut Tipi	5	540,062	0,004	1,195	0,312
	GPT Sürümü	1	20894,739	0,165	276,410	**<0,001***
	Malzeme Seçimi	1	56,279	<0,001	0,622	0,435
	GPT Sıcaklık Değeri	1	348,438	0,003	3,861	**0,046***
	En Beğenilen N Adet Tarif	1	53,679	<0,001	0,593	0,490
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	2	3585,343	0,028	20,371	**<0,001***
Toplam Jeton Adedi	Komut Tipi	5	4098697,244	0,111	34,971	**<0,001***
	GPT Sürümü	1	3058652,801	0,083	126,834	**<0,001***
	Malzeme Seçimi	1	185592,459	0,005	7,093	**0,012***
	GPT Sıcaklık Değeri	1	138099,325	0,004	5,271	**0,031***
	En Beğenilen N Adet Tarif	1	2648,460	<0,001	0,101	0,746
	LLM Tarafından Önerilmesi Beklenen Tarif Adedi	2	9962029,733	0,270	259,358	**<0,001***

EK 10 Deneysel sonuçlar (ilk iki koşum)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
1	1	top50_rand	gpt-4o	few_shot	1	0,2	3	0,2809	0,4803	1,3540	1,1667	9,16	1179
2	1	top50_rand	gpt-4o	zero_shot	1	0,2	3	0,5000	1,0000	1,0000	1,0000	4,50	971
3	1	top50_rand	gpt-4o	few_shot_role_play	1	0,2	3	0,5000	1,0000	1,0000	1,0000	5,68	1105
4	1	top50_rand	gpt-4o	few_shot_chain_of_thought	1	0,2	3	0,5000	1,0000	1,0000	1,0000	7,31	1153
5	1	top50_rand	gpt-4o	zero_shot_role_play	1	0,2	3	0,5000	1,0000	1,0000	1,0000	5,69	984
6	1	top50_rand	gpt-4o	zero_shot_chain_of_thought	1	0,2	3	0,5000	1,0000	1,0000	1,0000	4,73	1027
7	1	top50_rand	gpt-4o-mini	few_shot	1	0,2	3	1,0000	1,0000	0,0000	0,0000	5,19	1112
8	1	top50_rand	gpt-4o-mini	zero_shot	1	0,2	3	1,0000	1,0000	0,0000	0,0000	5,56	1104
9	1	top50_rand	gpt-4o-mini	few_shot_role_play	1	0,2	3	1,0000	1,0000	0,0000	0,0000	4,74	1167
10	1	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	1	0,2	3	1,0000	1,0000	0,0000	0,0000	9,02	1186
11	1	top50_rand	gpt-4o-mini	zero_shot_role_play	1	0,2	3	1,0000	1,0000	0,0000	0,0000	4,95	1001
12	1	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	1	0,2	3	1,0000	1,0000	0,0000	0,0000	4,45	1041
13	1	top50_rand	gpt-4o	few_shot	5	0,2	3	0,4376	0,6589	0,6325	0,4000	8,97	1379
14	1	top50_rand	gpt-4o	zero_shot	5	0,2	3	0,2360	0,3686	0,8944	0,8000	6,01	1114
15	1	top50_rand	gpt-4o	few_shot_role_play	5	0,2	3	0,6137	0,7355	0,6325	0,4000	9,86	1183
16	1	top50_rand	gpt-4o	few_shot_chain_of_thought	5	0,2	3	0,2940	0,4500	1,0954	0,8000	8,47	1502
17	1	top50_rand	gpt-4o	zero_shot_role_play	5	0,2	3	0,3500	0,4839	0,6325	0,4000	47,96	1266
18	1	top50_rand	gpt-4o	zero_shot_chain_of_thought	5	0,2	3	0,2937	0,4839	0,8944	0,8000	10,43	1368
19	1	top50_rand	gpt-4o-mini	few_shot	5	0,2	3	0,5561	0,5874	0,6325	0,4000	5,64	1226
20	1	top50_rand	gpt-4o-mini	zero_shot	5	0,2	3	0,5487	0,5874	1,0000	0,6000	6,48	1185
21	1	top50_rand	gpt-4o-mini	few_shot_role_play	5	0,2	3	0,5487	0,5874	1,0000	0,6000	7,17	1331
22	1	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	5	0,2	3	0,5404	0,5874	1,2649	0,8000	9,18	1416
23	1	top50_rand	gpt-4o-mini	zero_shot_role_play	5	0,2	3	0,4031	0,7355	0,8944	0,8000	7,59	1115
24	1	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	5	0,2	3	0,5561	0,5874	0,6325	0,4000	8,20	1188
25	1	top50_rand	gpt-4o	few_shot	10	0,2	3	0,4888	0,6388	1,0954	0,8000	9,93	1549
26	1	top50_rand	gpt-4o	zero_shot	10	0,2	3	0,3230	0,6177	1,0954	1,0000	10,59	1468
27	1	top50_rand	gpt-4o	few_shot_role_play	10	0,2	3	0,3907	0,5618	1,1402	0,9000	8,02	1468
28	1	top50_rand	gpt-4o	few_shot_chain_of_thought	10	0,2	3	0,2814	0,4987	1,0954	1,0000	10,14	1573
29	1	top50_rand	gpt-4o	zero_shot_role_play	10	0,2	3	0,6044	0,8188	0,6325	0,4000	19,97	1515
30	1	top50_rand	gpt-4o	zero_shot_chain_of_thought	10	0,2	3	0,4986	0,5988	0,7071	0,5000	8,70	1490
31	1	top50_rand	gpt-4o-mini	few_shot	10	0,2	3	0,4872	0,6745	1,0000	0,8000	11,23	1272
32	1	top50_rand	gpt-4o-mini	zero_shot	10	0,2	3	0,4677	0,6474	1,2649	1,0000	21,91	1161
33	1	top50_rand	gpt-4o-mini	few_shot_role_play	10	0,2	3	0,4677	0,6474	1,2649	1,0000	11,98	1313
34	1	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	10	0,2	3	0,4799	0,6388	1,5166	1,1000	18,75	1298
35	1	top50_rand	gpt-4o-mini	zero_shot_role_play	10	0,2	3	0,4863	0,6474	1,4142	1,0000	13,96	1214

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
36	1	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	10	0,2	3	0,4708	0,6474	1,3416	1,0000	13,04	1261
37	1	top50_rand	gpt-4o	few_shot	1	0,2	5	0,5000	1,0000	1,0000	1,0000	8,04	1401
38	1	top50_rand	gpt-4o	zero_shot	1	0,2	5	0,2602	0,5267	1,2649	1,2000	4,30	1035
39	1	top50_rand	gpt-4o	few_shot_role_play	1	0,2	5	0,5000	1,0000	1,0000	1,0000	4,55	1106
40	1	top50_rand	gpt-4o	few_shot_chain_of_thought	1	0,2	5	0,5000	1,0000	1,0000	1,0000	4,75	1114
41	1	top50_rand	gpt-4o	zero_shot_role_play	1	0,2	5	0,5000	1,0000	1,0000	1,0000	7,52	1362
42	1	top50_rand	gpt-4o	zero_shot_chain_of_thought	1	0,2	5	0,4839	0,4839	0,0000	0,0000	8,72	1327
43	1	top50_rand	gpt-4o-mini	few_shot	1	0,2	5	1,0000	1,0000	0,0000	0,0000	6,08	1135
44	1	top50_rand	gpt-4o-mini	zero_shot	1	0,2	5	1,0000	1,0000	0,0000	0,0000	5,34	972
45	1	top50_rand	gpt-4o-mini	few_shot_role_play	1	0,2	5	1,0000	1,0000	0,0000	0,0000	5,85	1100
46	1	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	1	0,2	5	1,0000	1,0000	0,0000	0,0000	4,88	1193
47	1	top50_rand	gpt-4o-mini	zero_shot_role_play	1	0,2	5	1,0000	1,0000	0,0000	0,0000	3,81	1076
48	1	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	1	0,2	5	1,0000	1,0000	0,0000	0,0000	4,36	1015
49	1	top50_rand	gpt-4o	few_shot	5	0,2	5	0,3455	0,5874	0,8944	0,8000	9,56	1353
50	1	top50_rand	gpt-4o	zero_shot	5	0,2	5	0,6624	0,7694	0,4472	0,2000	9,57	1367
51	1	top50_rand	gpt-4o	few_shot_role_play	5	0,2	5	0,6624	0,7694	0,4472	0,2000	8,18	1303
52	1	top50_rand	gpt-4o	few_shot_chain_of_thought	5	0,2	5	0,6624	0,7694	0,4472	0,2000	7,08	1338
53	1	top50_rand	gpt-4o	zero_shot_role_play	5	0,2	5	0,6070	1,0000	0,8944	0,8000	11,56	1184
54	1	top50_rand	gpt-4o	zero_shot_chain_of_thought	5	0,2	5	0,6213	0,7694	0,6325	0,4000	22,03	1241
55	1	top50_rand	gpt-4o-mini	few_shot	5	0,2	5	0,5487	0,5874	1,0000	0,6000	6,52	1360
56	1	top50_rand	gpt-4o-mini	zero_shot	5	0,2	5	0,6063	0,7355	1,0000	0,6000	11,61	1159
57	1	top50_rand	gpt-4o-mini	few_shot_role_play	5	0,2	5	0,5784	0,7355	0,7746	0,6000	7,73	1235
58	1	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	5	0,2	5	0,5784	0,7355	0,7746	0,6000	9,79	1300
59	1	top50_rand	gpt-4o-mini	zero_shot_role_play	5	0,2	5	0,5784	0,7355	0,7746	0,6000	6,16	1102
60	1	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	5	0,2	5	0,2833	0,4123	1,3416	1,0000	7,58	1437
61	1	top50_rand	gpt-4o	few_shot	10	0,2	5	0,2668	0,4849	1,2247	1,1000	8,23	1376
62	1	top50_rand	gpt-4o	zero_shot	10	0,2	5	0,5258	0,6177	0,6325	0,4000	7,60	1290
63	1	top50_rand	gpt-4o	few_shot_role_play	10	0,2	5	0,2893	0,5785	1,0000	1,0000	8,04	1417
64	1	top50_rand	gpt-4o	few_shot_chain_of_thought	10	0,2	5	0,3213	0,5618	1,1832	1,0000	7,48	1379
65	1	top50_rand	gpt-4o	zero_shot_role_play	10	0,2	5	0,5641	0,6474	0,8944	0,6000	8,18	1432
66	1	top50_rand	gpt-4o	zero_shot_chain_of_thought	10	0,2	5	0,5534	0,6745	0,7746	0,6000	11,42	1498
67	1	top50_rand	gpt-4o-mini	few_shot	10	0,2	5	0,5005	0,6663	1,1832	0,8000	10,10	1278
68	1	top50_rand	gpt-4o-mini	zero_shot	10	0,2	5	0,5034	0,6745	0,9487	0,7000	6,48	1253
69	1	top50_rand	gpt-4o-mini	few_shot_role_play	10	0,2	5	0,5005	0,6663	1,1832	0,8000	6,60	1305
70	1	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	10	0,2	5	0,4952	0,6663	1,0954	0,8000	6,76	1300
71	1	top50_rand	gpt-4o-mini	zero_shot_role_play	10	0,2	5	0,4710	0,6474	1,2247	0,9000	4,66	1161

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
72	1	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	10	0,2	5	0,4710	0,6474	1,2247	0,9000	5,75	1296
73	1	most_common	gpt-4o	few_shot	1	0,2	3	0,2419	0,4839	1,0000	1,0000	10,47	1463
74	1	most_common	gpt-4o	zero_shot	1	0,2	3	1,0000	1,0000	0,0000	0,0000	3,74	955
75	1	most_common	gpt-4o	few_shot_role_play	1	0,2	3	1,0000	1,0000	0,0000	0,0000	5,00	1135
76	1	most_common	gpt-4o	few_shot_chain_of_thought	1	0,2	3	1,0000	1,0000	0,0000	0,0000	12,48	1097
77	1	most_common	gpt-4o	zero_shot_role_play	1	0,2	3	1,0000	1,0000	0,0000	0,0000	30,89	1271
78	1	most_common	gpt-4o	zero_shot_chain_of_thought	1	0,2	3	1,0000	1,0000	0,0000	0,0000	25,35	1374
79	1	most_common	gpt-4o-mini	few_shot	1	0,2	3	0,5000	1,0000	1,0000	1,0000	5,99	1100
80	1	most_common	gpt-4o-mini	zero_shot	1	0,2	3	0,5000	1,0000	1,0000	1,0000	6,22	1056
81	1	most_common	gpt-4o-mini	few_shot_role_play	1	0,2	3	0,5000	1,0000	1,0000	1,0000	5,43	1166
82	1	most_common	gpt-4o-mini	few_shot_chain_of_thought	1	0,2	3	0,0313	1,0000	5,0000	5,0000	9,39	1730
83	1	most_common	gpt-4o-mini	zero_shot_role_play	1	0,2	3	0,5000	1,0000	1,0000	1,0000	4,42	1096
84	1	most_common	gpt-4o-mini	zero_shot_chain_of_thought	1	0,2	3	0,5000	1,0000	1,0000	1,0000	5,69	1071
85	1	most_common	gpt-4o	few_shot	5	0,2	3	0,6213	0,7694	0,6325	0,4000	11,51	1231
86	1	most_common	gpt-4o	zero_shot	5	0,2	3	0,6213	0,7694	0,6325	0,4000	8,59	1340
87	1	most_common	gpt-4o	few_shot_role_play	5	0,2	3	0,5860	0,7694	0,7746	0,6000	8,34	1399
88	1	most_common	gpt-4o	few_shot_chain_of_thought	5	0,2	3	0,5543	0,7694	0,8944	0,8000	10,88	1517
89	1	most_common	gpt-4o	zero_shot_role_play	5	0,2	3	0,6063	0,7355	1,0000	0,6000	11,40	1383
90	1	most_common	gpt-4o	zero_shot_chain_of_thought	5	0,2	3	0,6072	0,6589	0,4472	0,2000	11,19	1452
91	1	most_common	gpt-4o-mini	few_shot	5	0,2	3	0,3273	0,5874	1,1832	1,0000	16,87	1390
92	1	most_common	gpt-4o-mini	zero_shot	5	0,2	3	0,3815	0,6251	1,0954	0,8000	13,60	1195
93	1	most_common	gpt-4o-mini	few_shot_role_play	5	0,2	3	0,3265	0,5874	1,1832	1,0000	11,80	1289
94	1	most_common	gpt-4o-mini	few_shot_chain_of_thought	5	0,2	3	0,3265	0,5874	1,1832	1,0000	11,37	1283
95	1	most_common	gpt-4o-mini	zero_shot_role_play	5	0,2	3	0,5437	0,7355	1,0000	0,6000	12,52	1260
96	1	most_common	gpt-4o-mini	zero_shot_chain_of_thought	5	0,2	3	0,3273	0,5874	1,1832	1,0000	11,73	1232
97	1	most_common	gpt-4o	few_shot	10	0,2	3	0,5023	0,7095	1,5492	1,0000	7,66	1367
98	1	most_common	gpt-4o	zero_shot	10	0,2	3	0,1701	0,3468	1,7321	1,6000	8,84	1452
99	1	most_common	gpt-4o	few_shot_role_play	10	0,2	3	0,4334	0,5785	1,6125	1,2000	10,96	1487
100	1	most_common	gpt-4o	few_shot_chain_of_thought	10	0,2	3	0,4730	0,7749	2,3022	1,7000	33,41	1486
101	1	most_common	gpt-4o	zero_shot_role_play	10	0,2	3	0,5739	0,7259	0,7746	0,6000	31,08	1527
102	1	most_common	gpt-4o	zero_shot_chain_of_thought	10	0,2	3	0,5173	0,7259	1,9235	1,3000	17,18	1500
103	1	most_common	gpt-4o-mini	few_shot	10	0,2	3	0,3794	0,5618	1,9235	1,5000	14,15	1315
104	1	most_common	gpt-4o-mini	zero_shot	10	0,2	3	0,3969	0,5906	1,7029	1,3000	11,31	1184
105	1	most_common	gpt-4o-mini	few_shot_role_play	10	0,2	3	0,3025	0,5988	1,6733	1,4000	18,48	1342
106	1	most_common	gpt-4o-mini	few_shot_chain_of_thought	10	0,2	3	0,3975	0,5906	1,7607	1,3000	9,96	1295
107	1	most_common	gpt-4o-mini	zero_shot_role_play	10	0,2	3	0,3987	0,5988	1,6125	1,2000	11,50	1221

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
108	1	most_common	gpt-4o-mini	zero_shot_chain_of_thought	10	0,2	3	0,4247	0,6177	1,7321	1,2000	11,06	1222
109	1	most_common	gpt-4o	few_shot	1	0,2	5	0,5000	1,0000	1,0000	1,0000	16,75	1101
110	1	most_common	gpt-4o	zero_shot	1	0,2	5	1,0000	1,0000	0,0000	0,0000	13,90	971
111	1	most_common	gpt-4o	few_shot_role_play	1	0,2	5	1,0000	1,0000	0,0000	0,0000	30,87	1529
112	1	most_common	gpt-4o	few_shot_chain_of_thought	1	0,2	5	0,5000	1,0000	1,0000	1,0000	17,03	1140
113	1	most_common	gpt-4o	zero_shot_role_play	1	0,2	5	0,5000	1,0000	1,0000	1,0000	20,15	1069
114	1	most_common	gpt-4o	zero_shot_chain_of_thought	1	0,2	5	1,0000	1,0000	0,0000	0,0000	33,55	1819
115	1	most_common	gpt-4o-mini	few_shot	1	0,2	5	0,5000	1,0000	1,0000	1,0000	14,42	1155
116	1	most_common	gpt-4o-mini	zero_shot	1	0,2	5	0,5000	1,0000	1,0000	1,0000	9,63	1097
117	1	most_common	gpt-4o-mini	few_shot_role_play	1	0,2	5	0,5000	1,0000	1,0000	1,0000	7,96	1143
118	1	most_common	gpt-4o-mini	few_shot_chain_of_thought	1	0,2	5	0,5000	1,0000	1,0000	1,0000	9,53	1210
119	1	most_common	gpt-4o-mini	zero_shot_role_play	1	0,2	5	0,5000	1,0000	1,0000	1,0000	8,22	1074
120	1	most_common	gpt-4o-mini	zero_shot_chain_of_thought	1	0,2	5	0,5000	1,0000	1,0000	1,0000	6,22	1089
121	1	most_common	gpt-4o	few_shot	5	0,2	5	0,7499	0,8569	0,4472	0,2000	8,57	1322
122	1	most_common	gpt-4o	zero_shot	5	0,2	5	0,6213	0,7694	0,6325	0,4000	9,04	1306
123	1	most_common	gpt-4o	few_shot_role_play	5	0,2	5	0,3265	0,5874	1,1832	1,0000	8,98	1243
124	1	most_common	gpt-4o	few_shot_chain_of_thought	5	0,2	5	0,5653	0,7355	1,0954	0,8000	8,18	1268
125	1	most_common	gpt-4o	zero_shot_role_play	5	0,2	5	0,5710	0,7355	1,0954	0,8000	9,14	1258
126	1	most_common	gpt-4o	zero_shot_chain_of_thought	5	0,2	5	0,5478	0,5874	1,0000	0,6000	10,24	1326
127	1	most_common	gpt-4o-mini	few_shot	5	0,2	5	0,2854	0,5874	1,2649	1,2000	11,14	1282
128	1	most_common	gpt-4o-mini	zero_shot	5	0,2	5	0,2863	0,5874	1,2649	1,2000	10,06	1173
129	1	most_common	gpt-4o-mini	few_shot_role_play	5	0,2	5	0,2854	0,5874	1,2649	1,2000	8,92	1297
130	1	most_common	gpt-4o-mini	few_shot_chain_of_thought	5	0,2	5	0,2854	0,5874	1,2649	1,2000	8,75	1310
131	1	most_common	gpt-4o-mini	zero_shot_role_play	5	0,2	5	0,2863	0,5874	1,2649	1,2000	8,02	1219
132	1	most_common	gpt-4o-mini	zero_shot_chain_of_thought	5	0,2	5	0,3462	0,6251	1,1832	1,0000	10,24	1203
133	1	most_common	gpt-4o	few_shot	10	0,2	5	0,4125	0,5785	1,4832	1,2000	17,02	1345
134	1	most_common	gpt-4o	zero_shot	10	0,2	5	0,4990	0,7259	1,8974	1,2000	12,90	1558
135	1	most_common	gpt-4o	few_shot_role_play	10	0,2	5	0,5929	0,7749	1,8708	1,1000	11,94	1664
136	1	most_common	gpt-4o	few_shot_chain_of_thought	10	0,2	5	0,4250	0,5785	1,5492	1,2000	14,63	1554
137	1	most_common	gpt-4o	zero_shot_role_play	10	0,2	5	0,4432	0,5988	1,4491	1,1000	11,50	1580
138	1	most_common	gpt-4o	zero_shot_chain_of_thought	10	0,2	5	0,4252	0,5785	1,5492	1,2000	10,10	1447
139	1	most_common	gpt-4o-mini	few_shot	10	0,2	5	0,3993	0,5988	1,6125	1,2000	9,17	1283
140	1	most_common	gpt-4o-mini	zero_shot	10	0,2	5	0,4914	0,6745	1,4832	1,0000	9,71	1261
141	1	most_common	gpt-4o-mini	few_shot_role_play	10	0,2	5	0,3953	0,5906	1,4491	1,1000	8,65	1326
142	1	most_common	gpt-4o-mini	few_shot_chain_of_thought	10	0,2	5	0,4230	0,6177	1,5811	1,1000	10,79	1353
143	1	most_common	gpt-4o-mini	zero_shot_role_play	10	0,2	5	0,4914	0,6745	1,4832	1,0000	10,05	1241

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
144	1	most_common	gpt-4o-mini	zero_shot_chain_of_thought	10	0,2	5	0,4242	0,6356	1,7889	1,2000	13,32	1323
145	1	top50_rand	gpt-4o	few_shot	1	1,0	3	0,5000	1,0000	1,0000	1,0000	7,45	1151
146	1	top50_rand	gpt-4o	zero_shot	1	1,0	3	0,5000	1,0000	1,0000	1,0000	9,96	1136
147	1	top50_rand	gpt-4o	few_shot_role_play	1	1,0	3	0,2419	0,4839	1,0000	1,0000	6,23	1120
148	1	top50_rand	gpt-4o	few_shot_chain_of_thought	1	1,0	3	0,2419	0,4839	1,0000	1,0000	6,80	1106
149	1	top50_rand	gpt-4o	zero_shot_role_play	1	1,0	3	0,2419	0,4839	1,0000	1,0000	8,88	1130
150	1	top50_rand	gpt-4o	zero_shot_chain_of_thought	1	1,0	3	0,5000	1,0000	1,0000	1,0000	14,35	1345
151	1	top50_rand	gpt-4o-mini	few_shot	1	1,0	3	1,0000	1,0000	0,0000	0,0000	9,58	1175
152	1	top50_rand	gpt-4o-mini	zero_shot	1	1,0	3	1,0000	1,0000	0,0000	0,0000	7,41	1026
153	1	top50_rand	gpt-4o-mini	few_shot_role_play	1	1,0	3	0,5000	1,0000	1,0000	1,0000	7,75	1219
154	1	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	1	1,0	3	1,0000	1,0000	0,0000	0,0000	12,89	1224
155	1	top50_rand	gpt-4o-mini	zero_shot_role_play	1	1,0	3	1,0000	1,0000	0,0000	0,0000	6,00	1013
156	1	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	1	1,0	3	1,0000	1,0000	0,0000	0,0000	7,06	1035
157	1	top50_rand	gpt-4o	few_shot	5	1,0	3	0,6137	0,7355	0,6325	0,4000	8,73	1255
158	1	top50_rand	gpt-4o	zero_shot	5	1,0	3	0,2278	0,3686	1,1832	1,0000	8,64	1203
159	1	top50_rand	gpt-4o	few_shot_role_play	5	1,0	3	0,3191	0,5874	1,4142	1,2000	12,76	1731
160	1	top50_rand	gpt-4o	few_shot_chain_of_thought	5	1,0	3	0,7234	1,0000	0,6325	0,4000	12,38	1690
161	1	top50_rand	gpt-4o	zero_shot_role_play	5	1,0	3	0,4441	0,7355	0,7746	0,6000	9,80	1245
162	1	top50_rand	gpt-4o	zero_shot_chain_of_thought	5	1,0	3	0,5561	0,5874	0,6325	0,4000	13,83	1261
163	1	top50_rand	gpt-4o-mini	few_shot	5	1,0	3	0,4031	0,7355	0,8944	0,8000	12,42	1344
164	1	top50_rand	gpt-4o-mini	zero_shot	5	1,0	3	0,3603	0,7355	1,2649	1,2000	10,90	1183
165	1	top50_rand	gpt-4o-mini	few_shot_role_play	5	1,0	3	0,5561	0,5874	0,6325	0,4000	9,67	1268
166	1	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	5	1,0	3	0,5362	0,5704	1,2649	0,8000	11,73	1355
167	1	top50_rand	gpt-4o-mini	zero_shot_role_play	5	1,0	3	0,6651	0,8569	0,6325	0,4000	10,60	1222
168	1	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	5	1,0	3	0,6137	0,7355	0,6325	0,4000	13,10	1303
169	1	top50_rand	gpt-4o	few_shot	10	1,0	3	0,4888	0,6388	1,0954	0,8000	12,33	1426
170	1	top50_rand	gpt-4o	zero_shot	10	1,0	3	0,1591	0,2648	1,8439	1,6000	16,20	1381
171	1	top50_rand	gpt-4o	few_shot_role_play	10	1,0	3	0,5003	0,6556	0,7746	0,6000	15,44	1537
172	1	top50_rand	gpt-4o	few_shot_chain_of_thought	10	1,0	3	0,2256	0,4253	1,8439	1,6000	11,77	1454
173	1	top50_rand	gpt-4o	zero_shot_role_play	10	1,0	3	0,4426	0,5785	1,0000	0,8000	13,08	1327
174	1	top50_rand	gpt-4o	zero_shot_chain_of_thought	10	1,0	3	0,4695	0,5785	0,7746	0,6000	20,78	1370
175	1	top50_rand	gpt-4o-mini	few_shot	10	1,0	3	0,3626	0,4169	2,2804	1,8000	13,49	1273
176	1	top50_rand	gpt-4o-mini	zero_shot	10	1,0	3	0,5122	0,6924	0,9487	0,7000	11,14	1216
178	1	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	10	1,0	3	0,4631	0,6474	1,5492	1,2000	11,90	1367
179	1	top50_rand	gpt-4o-mini	zero_shot_role_play	10	1,0	3	0,3734	0,4169	2,0736	1,5000	10,44	1208
179	1	top50_rand	gpt-4o-mini	zero_shot_role_play	10	1,0	3	0,3734	0,4169	2,0736	1,5000	10,44	1208

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
181	1	top50_rand	gpt-4o	few_shot	1	1,0	5	0,5000	1,0000	1,0000	1,0000	6,88	1059
182	1	top50_rand	gpt-4o	zero_shot	1	1,0	5	0,5000	1,0000	1,0000	1,0000	10,18	1042
183	1	top50_rand	gpt-4o	few_shot_role_play	1	1,0	5	0,5000	1,0000	1,0000	1,0000	7,12	1142
184	1	top50_rand	gpt-4o	few_shot_chain_of_thought	1	1,0	5	0,5000	1,0000	1,0000	1,0000	6,59	1087
185	1	top50_rand	gpt-4o	zero_shot_role_play	1	1,0	5	0,5000	1,0000	1,0000	1,0000	6,62	987
186	1	top50_rand	gpt-4o	zero_shot_chain_of_thought	1	1,0	5	0,5000	1,0000	1,0000	1,0000	12,02	1298
187	1	top50_rand	gpt-4o-mini	few_shot	1	1,0	5	1,0000	1,0000	0,0000	0,0000	7,08	1096
188	1	top50_rand	gpt-4o-mini	zero_shot	1	1,0	5	0,4839	0,4839	0,0000	0,0000	12,69	1335
189	1	top50_rand	gpt-4o-mini	few_shot_role_play	1	1,0	5	1,0000	1,0000	0,0000	0,0000	5,97	1092
190	1	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	1	1,0	5	1,0000	1,0000	0,0000	0,0000	10,55	1250
191	1	top50_rand	gpt-4o-mini	zero_shot_role_play	1	1,0	5	1,0000	1,0000	0,0000	0,0000	7,08	1023
192	1	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	1	1,0	5	0,5000	1,0000	1,0000	1,0000	16,31	1560
193	1	top50_rand	gpt-4o	few_shot	5	1,0	5	0,2360	0,3686	0,8944	0,8000	13,24	1231
194	1	top50_rand	gpt-4o	zero_shot	5	1,0	5	0,6873	0,8569	0,4472	0,2000	9,87	1231
195	1	top50_rand	gpt-4o	few_shot_role_play	5	1,0	5	0,5511	0,7355	0,6325	0,4000	11,98	1366
196	1	top50_rand	gpt-4o	few_shot_chain_of_thought	5	1,0	5	0,6137	0,7355	0,6325	0,4000	10,30	1374
197	1	top50_rand	gpt-4o	zero_shot_role_play	5	1,0	5	0,6137	0,7355	0,6325	0,4000	7,35	1224
198	1	top50_rand	gpt-4o	zero_shot_chain_of_thought	5	1,0	5	0,3347	0,5874	0,8944	0,8000	15,38	1266
199	1	top50_rand	gpt-4o-mini	few_shot	5	1,0	5	0,5487	0,5874	1,0000	0,6000	9,07	1244
200	1	top50_rand	gpt-4o-mini	zero_shot	5	1,0	5	0,5561	0,5874	0,6325	0,4000	9,13	1072
201	1	top50_rand	gpt-4o-mini	few_shot_role_play	5	1,0	5	0,5478	0,5874	1,0000	0,6000	11,79	1363
202	1	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	5	1,0	5	0,6028	0,6251	0,8944	0,4000	15,81	1339
203	1	top50_rand	gpt-4o-mini	zero_shot_role_play	5	1,0	5	0,5362	0,5704	1,2649	0,8000	11,17	1188
204	1	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	5	1,0	5	0,3839	0,4884	1,6125	1,4000	7,15	1583
205	1	top50_rand	gpt-4o	few_shot	10	1,0	5	0,2930	0,5906	1,3416	1,2000	18,33	1362
206	1	top50_rand	gpt-4o	zero_shot	10	1,0	5	0,2247	0,3468	1,1832	1,0000	10,40	1243
207	1	top50_rand	gpt-4o	few_shot_role_play	10	1,0	5	0,3273	0,5618	1,1832	1,0000	15,33	1391
208	1	top50_rand	gpt-4o	few_shot_chain_of_thought	10	1,0	5	0,3657	0,5271	0,7746	0,6000	11,64	1448
209	1	top50_rand	gpt-4o	zero_shot_role_play	10	1,0	5	0,4580	0,5618	1,0954	0,8000	12,16	1424
210	1	top50_rand	gpt-4o	zero_shot_chain_of_thought	10	1,0	5	0,2925	0,5069	1,0488	0,9000	11,98	1457
211	1	top50_rand	gpt-4o-mini	few_shot	10	1,0	5	0,4366	0,5618	1,3416	1,0000	10,57	1329
212	1	top50_rand	gpt-4o-mini	zero_shot	10	1,0	5	0,4673	0,5820	1,3038	0,9000	10,98	1344
213	1	top50_rand	gpt-4o-mini	few_shot_role_play	10	1,0	5	0,5010	0,6745	0,9487	0,7000	8,13	1385
214	1	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	10	1,0	5	0,4796	0,5988	0,7746	0,6000	9,31	1325
216	1	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	10	1,0	5	0,4427	0,6271	1,0488	0,9000	9,85	1235
217	1	most_common	gpt-4o	few_shot	1	1,0	3	0,5000	1,0000	1,0000	1,0000	13,94	1652

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
218	1	most_common	gpt-4o	zero_shot	1	1,0	3	1,0000	1,0000	0,0000	0,0000	16,86	1693
219	1	most_common	gpt-4o	few_shot_role_play	1	1,0	3	1,0000	1,0000	0,0000	0,0000	7,13	1164
220	1	most_common	gpt-4o	few_shot_chain_of_thought	1	1,0	3	0,2419	0,4839	1,0000	1,0000	7,00	1228
221	1	most_common	gpt-4o	zero_shot_role_play	1	1,0	3	1,0000	1,0000	0,0000	0,0000	14,60	1218
222	1	most_common	gpt-4o	zero_shot_chain_of_thought	1	1,0	3	1,0000	1,0000	0,0000	0,0000	25,22	1355
223	1	most_common	gpt-4o-mini	few_shot	1	1,0	3	0,5000	1,0000	1,0000	1,0000	6,96	1167
224	1	most_common	gpt-4o-mini	zero_shot	1	1,0	3	0,5000	1,0000	1,0000	1,0000	5,47	1046
225	1	most_common	gpt-4o-mini	few_shot_role_play	1	1,0	3	0,5000	1,0000	1,0000	1,0000	7,26	1098
226	1	most_common	gpt-4o-mini	few_shot_chain_of_thought	1	1,0	3	1,0000	1,0000	0,0000	0,0000	16,01	1595
227	1	most_common	gpt-4o-mini	zero_shot_role_play	1	1,0	3	0,5000	1,0000	1,0000	1,0000	8,07	1119
228	1	most_common	gpt-4o-mini	zero_shot_chain_of_thought	1	1,0	3	0,5000	1,0000	1,0000	1,0000	5,61	1090
229	1	most_common	gpt-4o	few_shot	5	1,0	3	0,5543	0,7694	0,8944	0,8000	10,98	1594
230	1	most_common	gpt-4o	zero_shot	5	1,0	3	0,6624	0,7694	0,4472	0,2000	8,40	1287
231	1	most_common	gpt-4o	few_shot_role_play	5	1,0	3	0,5540	0,7853	1,1832	1,0000	9,17	1435
232	1	most_common	gpt-4o	few_shot_chain_of_thought	5	1,0	3	0,6028	0,6251	0,8944	0,4000	10,19	1319
233	1	most_common	gpt-4o	zero_shot_role_play	5	1,0	3	0,5511	0,6251	1,0000	0,6000	8,59	1193
234	1	most_common	gpt-4o	zero_shot_chain_of_thought	5	1,0	3	0,5076	0,5874	1,0954	0,8000	13,95	1335
235	1	most_common	gpt-4o-mini	few_shot	5	1,0	3	0,2937	0,5874	1,0000	1,0000	8,05	1203
236	1	most_common	gpt-4o-mini	zero_shot	5	1,0	3	0,2854	0,5874	1,2649	1,2000	98,91	1161
237	1	most_common	gpt-4o-mini	few_shot_role_play	5	1,0	3	0,3265	0,5874	1,1832	1,0000	9,53	1268
238	1	most_common	gpt-4o-mini	few_shot_chain_of_thought	5	1,0	3	0,3922	0,6251	1,0954	0,8000	8,74	1248
239	1	most_common	gpt-4o-mini	zero_shot_role_play	5	1,0	3	0,2854	0,5874	1,2649	1,2000	11,18	1257
240	1	most_common	gpt-4o-mini	zero_shot_chain_of_thought	5	1,0	3	0,3957	0,7355	1,1832	1,0000	11,62	1302
242	1	most_common	gpt-4o	zero_shot	10	1,0	3	0,4119	0,5398	1,7889	1,4000	9,80	1416
244	1	most_common	gpt-4o	few_shot_chain_of_thought	10	1,0	3	0,4924	0,7259	1,4832	1,0000	29,89	1898
245	1	most_common	gpt-4o	zero_shot_role_play	10	1,0	3	0,4837	0,6527	1,4491	0,9000	25,59	1487
246	1	most_common	gpt-4o	zero_shot_chain_of_thought	10	1,0	3	0,2474	0,5618	2,4698	1,9000	16,39	1614
247	1	most_common	gpt-4o-mini	few_shot	10	1,0	3	0,3959	0,5820	1,7029	1,3000	12,33	1315
248	1	most_common	gpt-4o-mini	zero_shot	10	1,0	3	0,2861	0,5618	1,8974	1,6000	10,28	1212
249	1	most_common	gpt-4o-mini	few_shot_role_play	10	1,0	3	0,3063	0,6177	1,7321	1,4000	10,25	1306
250	1	most_common	gpt-4o-mini	few_shot_chain_of_thought	10	1,0	3	0,2700	0,5988	1,8974	1,6000	11,00	1377
251	1	most_common	gpt-4o-mini	zero_shot_role_play	10	1,0	3	0,2868	0,5906	1,8166	1,5000	10,48	1241
252	1	most_common	gpt-4o-mini	zero_shot_chain_of_thought	10	1,0	3	0,2702	0,4681	1,7321	1,4000	9,35	1214
253	1	most_common	gpt-4o	few_shot	1	1,0	5	1,0000	1,0000	0,0000	0,0000	11,89	1293
254	1	most_common	gpt-4o	zero_shot	1	1,0	5	0,2419	0,4839	1,0000	1,0000	11,26	1215
255	1	most_common	gpt-4o	few_shot_role_play	1	1,0	5	0,5000	1,0000	1,0000	1,0000	5,62	1081

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
256	1	most_common	gpt-4o	few_shot_chain_of_thought	1	1,0	5	0,5000	1,0000	1,0000	1,0000	14,59	1393
257	1	most_common	gpt-4o	zero_shot_role_play	1	1,0	5	0,5000	1,0000	1,0000	1,0000	8,70	1004
258	1	most_common	gpt-4o	zero_shot_chain_of_thought	1	1,0	5	1,0000	1,0000	0,0000	0,0000	14,08	1243
259	1	most_common	gpt-4o-mini	few_shot	1	1,0	5	0,5000	1,0000	1,0000	1,0000	7,86	1138
260	1	most_common	gpt-4o-mini	zero_shot	1	1,0	5	0,5000	1,0000	1,0000	1,0000	8,06	1035
261	1	most_common	gpt-4o-mini	few_shot_role_play	1	1,0	5	0,5000	1,0000	1,0000	1,0000	8,06	1186
262	1	most_common	gpt-4o-mini	few_shot_chain_of_thought	1	1,0	5	0,5000	1,0000	1,0000	1,0000	10,08	1275
263	1	most_common	gpt-4o-mini	zero_shot_role_play	1	1,0	5	0,5000	1,0000	1,0000	1,0000	11,80	1136
264	1	most_common	gpt-4o-mini	zero_shot_chain_of_thought	1	1,0	5	0,5000	1,0000	1,0000	1,0000	10,33	1124
265	1	most_common	gpt-4o	few_shot	5	1,0	5	0,5084	0,7355	1,0954	0,8000	19,81	1536
266	1	most_common	gpt-4o	zero_shot	5	1,0	5	0,5860	0,7694	0,7746	0,6000	16,81	1344
267	1	most_common	gpt-4o	few_shot_role_play	5	1,0	5	0,3220	0,6251	1,1832	1,0000	12,88	1418
268	1	most_common	gpt-4o	few_shot_chain_of_thought	5	1,0	5	0,4957	0,5874	1,6733	1,2000	12,54	1377
269	1	most_common	gpt-4o	zero_shot_role_play	5	1,0	5	0,5826	0,6589	0,6325	0,4000	11,33	1302
270	1	most_common	gpt-4o	zero_shot_chain_of_thought	5	1,0	5	0,5299	0,7355	1,1832	1,0000	9,05	1253
271	1	most_common	gpt-4o-mini	few_shot	5	1,0	5	0,3462	0,6251	1,1832	1,0000	8,28	1259
272	1	most_common	gpt-4o-mini	zero_shot	5	1,0	5	0,2780	0,5874	1,4832	1,4000	9,44	1232
273	1	most_common	gpt-4o-mini	few_shot_role_play	5	1,0	5	0,5027	0,7355	1,0954	0,8000	7,68	1258
274	1	most_common	gpt-4o-mini	few_shot_chain_of_thought	5	1,0	5	0,2780	0,5874	1,4832	1,4000	6,14	1275
275	1	most_common	gpt-4o-mini	zero_shot_role_play	5	1,0	5	0,5234	0,7694	0,7746	0,6000	7,98	1147
276	1	most_common	gpt-4o-mini	zero_shot_chain_of_thought	5	1,0	5	0,2738	0,5704	1,4832	1,4000	7,12	1259
277	1	most_common	gpt-4o	few_shot	10	1,0	5	0,4395	0,5988	1,5492	1,2000	14,45	1500
278	1	most_common	gpt-4o	zero_shot	10	1,0	5	0,4867	0,6691	1,4491	0,9000	12,66	1559
278	1	most_common	gpt-4o	zero_shot	10	1,0	5	0,4867	0,6691	1,4491	0,9000	12,66	1559
280	1	most_common	gpt-4o	few_shot_chain_of_thought	10	1,0	5	0,4556	0,6924	1,6125	1,2000	12,84	1449
281	1	most_common	gpt-4o	zero_shot_role_play	10	1,0	5	0,4519	0,6177	1,0000	0,8000	12,54	1412
282	1	most_common	gpt-4o	zero_shot_chain_of_thought	10	1,0	5	0,3928	0,5069	1,4491	1,1000	11,98	1303
283	1	most_common	gpt-4o-mini	few_shot	10	1,0	5	0,4190	0,6745	1,2649	1,0000	9,26	1349
284	1	most_common	gpt-4o-mini	zero_shot	10	1,0	5	0,4951	0,6745	1,0954	0,8000	8,40	1228
285	1	most_common	gpt-4o-mini	few_shot_role_play	10	1,0	5	0,5263	0,6924	1,1402	0,7000	9,41	1418
286	1	most_common	gpt-4o-mini	few_shot_chain_of_thought	10	1,0	5	0,3898	0,5906	1,0488	0,9000	8,33	1396
287	1	most_common	gpt-4o-mini	zero_shot_role_play	10	1,0	5	0,3208	0,5378	1,1402	0,9000	9,24	1209
288	1	most_common	gpt-4o-mini	zero_shot_chain_of_thought	10	1,0	5	0,4496	0,6177	1,1402	0,9000	10,13	1470
1	2	top50_rand	gpt-4o	few_shot	1	0,2	3	0,1567	0,3469	1,4142	1,3333	10,32	1245
2	2	top50_rand	gpt-4o	zero_shot	1	0,2	3	0,2419	0,4839	1,0000	1,0000	11,15	1000
3	2	top50_rand	gpt-4o	few_shot_role_play	1	0,2	3	1,0000	1,0000	0,0000	0,0000	9,28	1180

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
4	2	top50_rand	gpt-4o	few_shot_chain_of_thought	1	0,2	3	0,2419	0,4839	1,0000	1,0000	13,36	1167
5	2	top50_rand	gpt-4o	zero_shot_role_play	1	0,2	3	0,5000	1,0000	1,0000	1,0000	12,42	1074
6	2	top50_rand	gpt-4o	zero_shot_chain_of_thought	1	0,2	3	0,4839	0,4839	0,0000	0,0000	7,95	1080
7	2	top50_rand	gpt-4o-mini	few_shot	1	0,2	3	0,5000	1,0000	1,0000	1,0000	6,09	1138
8	2	top50_rand	gpt-4o-mini	zero_shot	1	0,2	3	0,5000	1,0000	1,0000	1,0000	6,55	1131
9	2	top50_rand	gpt-4o-mini	few_shot_role_play	1	0,2	3	0,5000	1,0000	1,0000	1,0000	7,64	1193
10	2	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	1	0,2	3	0,5000	1,0000	1,0000	1,0000	6,08	1165
11	2	top50_rand	gpt-4o-mini	zero_shot_role_play	1	0,2	3	0,5000	1,0000	1,0000	1,0000	6,92	1130
12	2	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	1	0,2	3	0,5000	1,0000	1,0000	1,0000	7,05	1107
13	2	top50_rand	gpt-4o	few_shot	5	0,2	3	0,2737	0,4839	0,8944	0,8000	16,97	1665
14	2	top50_rand	gpt-4o	zero_shot	5	0,2	3	0,5308	0,6589	0,7746	0,6000	18,82	1596
15	2	top50_rand	gpt-4o	few_shot_role_play	5	0,2	3	0,2419	0,4839	1,0000	1,0000	28,93	1794
16	2	top50_rand	gpt-4o	few_shot_chain_of_thought	5	0,2	3	0,2737	0,4839	0,8944	0,8000	35,38	1676
17	2	top50_rand	gpt-4o	zero_shot_role_play	5	0,2	3	0,7384	0,7853	1,2649	0,8000	18,64	1637
18	2	top50_rand	gpt-4o	zero_shot_chain_of_thought	5	0,2	3	0,2937	0,5874	1,0000	1,0000	12,15	1276
19	2	top50_rand	gpt-4o-mini	few_shot	5	0,2	3	0,2854	0,5874	1,2649	1,2000	12,69	1360
20	2	top50_rand	gpt-4o-mini	zero_shot	5	0,2	3	0,2863	0,5874	1,2649	1,2000	10,60	1313
21	2	top50_rand	gpt-4o-mini	few_shot_role_play	5	0,2	3	0,4960	0,5874	1,0954	0,8000	8,88	1323
22	2	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	5	0,2	3	0,2907	0,4123	1,0954	0,8000	9,08	1448
23	2	top50_rand	gpt-4o-mini	zero_shot_role_play	5	0,2	3	0,2863	0,5874	1,2649	1,2000	7,63	1243
24	2	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	5	0,2	3	0,2937	0,5874	1,0000	1,0000	6,15	1211
25	2	top50_rand	gpt-4o	few_shot	10	0,2	3	0,2252	0,4849	1,4832	1,4000	11,67	1437
26	2	top50_rand	gpt-4o	zero_shot	10	0,2	3	0,3678	0,4849	1,1832	1,0000	14,66	1446
27	2	top50_rand	gpt-4o	few_shot_role_play	10	0,2	3	0,1927	0,3468	1,2247	1,1000	13,63	1531
28	2	top50_rand	gpt-4o	few_shot_chain_of_thought	10	0,2	3	0,1272	0,2833	1,8708	1,7000	12,61	1519
29	2	top50_rand	gpt-4o	zero_shot_role_play	10	0,2	3	0,3635	0,5069	0,9487	0,9000	13,08	1417
30	2	top50_rand	gpt-4o	zero_shot_chain_of_thought	10	0,2	3	0,2678	0,5069	1,3416	1,2000	14,59	1502
31	2	top50_rand	gpt-4o-mini	few_shot	10	0,2	3	0,2426	0,4987	1,4491	1,3000	8,84	1325
32	2	top50_rand	gpt-4o-mini	zero_shot	10	0,2	3	0,2467	0,4681	1,7607	1,5000	7,02	1218
33	2	top50_rand	gpt-4o-mini	few_shot_role_play	10	0,2	3	0,2763	0,5189	1,4142	1,2000	8,41	1327
34	2	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	10	0,2	3	0,2426	0,4987	1,4491	1,3000	9,05	1403
35	2	top50_rand	gpt-4o-mini	zero_shot_role_play	10	0,2	3	0,2618	0,4987	1,5166	1,3000	8,82	1269
36	2	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	10	0,2	3	0,2467	0,4681	1,7607	1,5000	11,23	1303
37	2	top50_rand	gpt-4o	few_shot	1	0,2	5	1,0000	1,0000	0,0000	0,0000	7,78	1052
38	2	top50_rand	gpt-4o	zero_shot	1	0,2	5	1,0000	1,0000	0,0000	0,0000	7,53	1018
39	2	top50_rand	gpt-4o	few_shot_role_play	1	0,2	5	1,0000	1,0000	0,0000	0,0000	20,70	1185

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
40	2	top50_rand	gpt-4o	few_shot_chain_of_thought	1	0,2	5	1,0000	1,0000	0,0000	0,0000	24,24	1158
41	2	top50_rand	gpt-4o	zero_shot_role_play	1	0,2	5	1,0000	1,0000	0,0000	0,0000	32,38	1043
43	2	top50_rand	gpt-4o-mini	few_shot	1	0,2	5	0,4839	0,4839	0,0000	0,0000	9,94	1430
44	2	top50_rand	gpt-4o-mini	zero_shot	1	0,2	5	0,2419	0,4839	1,0000	1,0000	5,59	1067
45	2	top50_rand	gpt-4o-mini	few_shot_role_play	1	0,2	5	0,2419	0,4839	1,0000	1,0000	7,94	1220
46	2	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	1	0,2	5	0,2419	0,4839	1,0000	1,0000	6,15	1175
47	2	top50_rand	gpt-4o-mini	zero_shot_role_play	1	0,2	5	0,2419	0,4839	1,0000	1,0000	7,04	1147
48	2	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	1	0,2	5	0,2419	0,4839	1,0000	1,0000	6,88	1082
49	2	top50_rand	gpt-4o	few_shot	5	0,2	5	0,1194	0,3133	1,8439	1,8000	16,97	1562
50	2	top50_rand	gpt-4o	zero_shot	5	0,2	5	0,2737	0,4839	0,8944	0,8000	23,03	1406
51	2	top50_rand	gpt-4o	few_shot_role_play	5	0,2	5	0,6966	0,7694	0,6325	0,4000	15,15	1494
55	2	top50_rand	gpt-4o-mini	few_shot	5	0,2	5	0,2066	0,3516	1,6125	1,4000	19,22	1300
56	2	top50_rand	gpt-4o-mini	zero_shot	5	0,2	5	0,2148	0,3516	1,4142	1,2000	10,32	1197
57	2	top50_rand	gpt-4o-mini	few_shot_role_play	5	0,2	5	0,4994	0,5874	1,3416	1,0000	9,85	1339
58	2	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	5	0,2	5	0,2108	0,3686	1,6125	1,4000	8,99	1324
59	2	top50_rand	gpt-4o-mini	zero_shot_role_play	5	0,2	5	0,4994	0,5874	1,3416	1,0000	8,96	1221
60	2	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	5	0,2	5	0,4994	0,5874	1,3416	1,0000	10,00	1317
61	2	top50_rand	gpt-4o	few_shot	10	0,2	5	0,4476	0,5810	0,9487	0,7000	57,16	1813
62	2	top50_rand	gpt-4o	zero_shot	10	0,2	5	0,5941	0,6556	1,0488	0,7000	12,98	1307
63	2	top50_rand	gpt-4o	few_shot_role_play	10	0,2	5	0,5106	0,5785	1,0954	0,8000	17,18	1371
64	2	top50_rand	gpt-4o	few_shot_chain_of_thought	10	0,2	5	0,3433	0,4849	1,2247	1,1000	20,90	1503
65	2	top50_rand	gpt-4o	zero_shot_role_play	10	0,2	5	0,3539	0,4681	1,5811	1,3000	14,55	1292
65	2	top50_rand	gpt-4o	zero_shot_role_play	10	0,2	5	0,3539	0,4681	1,5811	1,3000	14,55	1292
67	2	top50_rand	gpt-4o-mini	few_shot	10	0,2	5	0,3943	0,5618	1,8166	1,5000	11,94	1379
68	2	top50_rand	gpt-4o-mini	zero_shot	10	0,2	5	0,4216	0,5906	1,4832	1,2000	10,38	1172
69	2	top50_rand	gpt-4o-mini	few_shot_role_play	10	0,2	5	0,4380	0,6177	1,1832	1,0000	11,13	1348
70	2	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	10	0,2	5	0,4195	0,5906	1,3784	1,1000	12,12	1391
71	2	top50_rand	gpt-4o-mini	zero_shot_role_play	10	0,2	5	0,4293	0,6177	1,1832	1,0000	19,52	1240
72	2	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	10	0,2	5	0,4331	0,6177	1,0488	0,9000	9,32	1228
73	2	most_common	gpt-4o	few_shot	1	0,2	3	0,5000	1,0000	1,0000	1,0000	8,16	1116
74	2	most_common	gpt-4o	zero_shot	1	0,2	3	0,5000	1,0000	1,0000	1,0000	8,09	1032
75	2	most_common	gpt-4o	few_shot_role_play	1	0,2	3	0,5000	1,0000	1,0000	1,0000	10,53	1147
76	2	most_common	gpt-4o	few_shot_chain_of_thought	1	0,2	3	0,5000	1,0000	1,0000	1,0000	11,32	1148
77	2	most_common	gpt-4o	zero_shot_role_play	1	0,2	3	0,5000	1,0000	1,0000	1,0000	8,30	1037
78	2	most_common	gpt-4o	zero_shot_chain_of_thought	1	0,2	3	1,0000	1,0000	0,0000	0,0000	15,67	1028
79	2	most_common	gpt-4o-mini	few_shot	1	0,2	3	0,5000	1,0000	1,0000	1,0000	8,24	1144

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
80	2	most_common	gpt-4o-mini	zero_shot	1	0,2	3	0,5000	1,0000	1,0000	1,0000	15,05	1004
81	2	most_common	gpt-4o-mini	few_shot_role_play	1	0,2	3	0,5000	1,0000	1,0000	1,0000	7,42	1140
82	2	most_common	gpt-4o-mini	few_shot_chain_of_thought	1	0,2	3	0,5000	1,0000	1,0000	1,0000	8,35	1157
83	2	most_common	gpt-4o-mini	zero_shot_role_play	1	0,2	3	0,5000	1,0000	1,0000	1,0000	7,57	986
84	2	most_common	gpt-4o-mini	zero_shot_chain_of_thought	1	0,2	3	0,5000	1,0000	1,0000	1,0000	12,85	1524
85	2	most_common	gpt-4o	few_shot	5	0,2	3	0,4886	0,5874	1,3416	1,0000	27,08	1383
86	2	most_common	gpt-4o	zero_shot	5	0,2	3	0,2940	0,4500	1,0954	0,8000	15,33	1390
87	2	most_common	gpt-4o	few_shot_role_play	5	0,2	3	0,5754	0,6589	0,6325	0,4000	12,23	1421
88	2	most_common	gpt-4o	few_shot_chain_of_thought	5	0,2	3	0,5478	0,5874	1,0000	0,6000	17,75	1454
89	2	most_common	gpt-4o	zero_shot_role_play	5	0,2	3	0,2940	0,4500	1,0954	0,8000	19,35	1438
90	2	most_common	gpt-4o	zero_shot_chain_of_thought	5	0,2	3	0,7050	0,8569	0,7746	0,6000	17,22	1559
91	2	most_common	gpt-4o-mini	few_shot	5	0,2	3	0,3265	0,5874	1,1832	1,0000	8,95	1236
92	2	most_common	gpt-4o-mini	zero_shot	5	0,2	3	0,3265	0,5874	1,1832	1,0000	7,11	1132
93	2	most_common	gpt-4o-mini	few_shot_role_play	5	0,2	3	0,3265	0,5874	1,1832	1,0000	7,79	1284
94	2	most_common	gpt-4o-mini	few_shot_chain_of_thought	5	0,2	3	0,3265	0,5874	1,1832	1,0000	10,20	1277
95	2	most_common	gpt-4o-mini	zero_shot_role_play	5	0,2	3	0,3372	0,5874	1,1832	1,0000	8,52	1224
96	2	most_common	gpt-4o-mini	zero_shot_chain_of_thought	5	0,2	3	0,3372	0,5874	1,1832	1,0000	7,86	1185
97	2	most_common	gpt-4o	few_shot	10	0,2	3	0,5106	0,6691	0,7071	0,5000	29,88	1580
98	2	most_common	gpt-4o	zero_shot	10	0,2	3	0,5543	0,7694	0,8944	0,8000	20,69	1620
100	2	most_common	gpt-4o	few_shot_chain_of_thought	10	0,2	3	0,4651	0,6527	1,9748	1,3000	24,68	1480
102	2	most_common	gpt-4o	zero_shot_chain_of_thought	10	0,2	3	0,1701	0,3468	1,7321	1,6000	18,00	1522
103	2	most_common	gpt-4o-mini	few_shot	10	0,2	3	0,3233	0,5988	1,6432	1,3000	9,09	1286
104	2	most_common	gpt-4o-mini	zero_shot	10	0,2	3	0,3038	0,5906	1,8974	1,6000	7,70	1122
105	2	most_common	gpt-4o-mini	few_shot_role_play	10	0,2	3	0,3969	0,5906	1,7029	1,3000	8,95	1249
106	2	most_common	gpt-4o-mini	few_shot_chain_of_thought	10	0,2	3	0,3771	0,5618	1,9235	1,5000	8,94	1361
107	2	most_common	gpt-4o-mini	zero_shot_role_play	10	0,2	3	0,3946	0,5906	1,7029	1,3000	7,15	1174
108	2	most_common	gpt-4o-mini	zero_shot_chain_of_thought	10	0,2	3	0,3666	0,5906	1,9748	1,5000	7,67	1211
109	2	most_common	gpt-4o	few_shot	1	0,2	5	0,5000	1,0000	1,0000	1,0000	6,95	1124
110	2	most_common	gpt-4o	zero_shot	1	0,2	5	0,5000	1,0000	1,0000	1,0000	6,71	972
111	2	most_common	gpt-4o	few_shot_role_play	1	0,2	5	0,5000	1,0000	1,0000	1,0000	8,73	1121
112	2	most_common	gpt-4o	few_shot_chain_of_thought	1	0,2	5	0,5000	1,0000	1,0000	1,0000	15,95	1137
113	2	most_common	gpt-4o	zero_shot_role_play	1	0,2	5	1,0000	1,0000	0,0000	0,0000	8,06	1044
114	2	most_common	gpt-4o	zero_shot_chain_of_thought	1	0,2	5	1,0000	1,0000	0,0000	0,0000	9,08	1063
115	2	most_common	gpt-4o-mini	few_shot	1	0,2	5	0,5000	1,0000	1,0000	1,0000	6,44	1158
116	2	most_common	gpt-4o-mini	zero_shot	1	0,2	5	0,5000	1,0000	1,0000	1,0000	6,68	1045
117	2	most_common	gpt-4o-mini	few_shot_role_play	1	0,2	5	0,5000	1,0000	1,0000	1,0000	7,01	1170

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
118	2	most_common	gpt-4o-mini	few_shot_chain_of_thought	1	0,2	5	0,5000	1,0000	1,0000	1,0000	6,93	1185
119	2	most_common	gpt-4o-mini	zero_shot_role_play	1	0,2	5	0,5000	1,0000	1,0000	1,0000	6,79	1083
120	2	most_common	gpt-4o-mini	zero_shot_chain_of_thought	1	0,2	5	0,5000	1,0000	1,0000	1,0000	7,20	1099
121	2	most_common	gpt-4o	few_shot	5	0,2	5	0,6306	0,7694	0,6325	0,4000	14,51	1226
122	2	most_common	gpt-4o	zero_shot	5	0,2	5	0,5953	0,7694	0,7746	0,6000	13,44	1100
123	2	most_common	gpt-4o	few_shot_role_play	5	0,2	5	0,6270	0,7694	0,6325	0,4000	10,75	1191
124	2	most_common	gpt-4o	few_shot_chain_of_thought	5	0,2	5	0,6270	0,7694	0,6325	0,4000	30,23	1295
125	2	most_common	gpt-4o	zero_shot_role_play	5	0,2	5	0,5710	0,7355	1,0954	0,8000	8,85	1143
126	2	most_common	gpt-4o	zero_shot_chain_of_thought	5	0,2	5	0,4775	0,5436	1,3416	1,0000	13,64	1413
127	2	most_common	gpt-4o-mini	few_shot	5	0,2	5	0,2937	0,5874	1,0000	1,0000	8,52	1301
128	2	most_common	gpt-4o-mini	zero_shot	5	0,2	5	0,2854	0,5874	1,2649	1,2000	8,99	1142
129	2	most_common	gpt-4o-mini	few_shot_role_play	5	0,2	5	0,2854	0,5874	1,2649	1,2000	8,45	1296
130	2	most_common	gpt-4o-mini	few_shot_chain_of_thought	5	0,2	5	0,2780	0,5874	1,4832	1,4000	8,94	1383
131	2	most_common	gpt-4o-mini	zero_shot_role_play	5	0,2	5	0,2863	0,5874	1,2649	1,2000	11,26	1256
132	2	most_common	gpt-4o-mini	zero_shot_chain_of_thought	5	0,2	5	0,2863	0,5874	1,2649	1,2000	7,30	1221
133	2	most_common	gpt-4o	few_shot	10	0,2	5	0,5543	0,7694	0,8944	0,8000	34,13	1628
134	2	most_common	gpt-4o	zero_shot	10	0,2	5	0,4884	0,7259	0,8944	0,8000	18,98	1466
135	2	most_common	gpt-4o	few_shot_role_play	10	0,2	5	0,3705	0,4604	1,6733	1,4000	21,24	1763
135	2	most_common	gpt-4o	few_shot_role_play	10	0,2	5	0,4938	0,6691	1,8708	1,1000	27,09	1704
136	2	most_common	gpt-4o	few_shot_chain_of_thought	10	0,2	5	0,5101	0,7259	1,4491	0,9000	27,47	1672
136	2	most_common	gpt-4o	few_shot_chain_of_thought	10	0,2	5	0,3644	0,4849	1,5492	1,2000	25,16	1622
137	2	most_common	gpt-4o	zero_shot_role_play	10	0,2	5	0,4675	0,7259	2,2804	1,6000	32,31	1554
139	2	most_common	gpt-4o-mini	few_shot	10	0,2	5	0,3808	0,5988	1,5492	1,2000	9,90	1328
138	2	most_common	gpt-4o	zero_shot_chain_of_thought	10	0,2	5	0,6038	0,8593	1,7321	1,0000	61,50	1505
140	2	most_common	gpt-4o-mini	zero_shot	10	0,2	5	0,3956	0,5988	1,6125	1,2000	17,69	1175
141	2	most_common	gpt-4o-mini	few_shot_role_play	10	0,2	5	0,3808	0,5988	1,5492	1,2000	7,46	1273
140	2	most_common	gpt-4o-mini	zero_shot	10	0,2	5	0,3664	0,5618	1,6432	1,3000	6,50	1112
142	2	most_common	gpt-4o-mini	few_shot_chain_of_thought	10	0,2	5	0,4872	0,6745	1,4832	1,0000	7,74	1329
143	2	most_common	gpt-4o-mini	zero_shot_role_play	10	0,2	5	0,3763	0,5906	1,3784	1,1000	9,04	1135
144	2	most_common	gpt-4o-mini	zero_shot_chain_of_thought	10	0,2	5	0,4942	0,6745	1,4832	1,0000	10,02	1249
145	2	top50_rand	gpt-4o	few_shot	1	1,0	3	1,0000	1,0000	0,0000	0,0000	9,24	1176
147	2	top50_rand	gpt-4o	few_shot_role_play	1	1,0	3	0,5000	1,0000	1,0000	1,0000	7,06	1169
148	2	top50_rand	gpt-4o	few_shot_chain_of_thought	1	1,0	3	0,2419	0,4839	1,0000	1,0000	15,32	1245
149	2	top50_rand	gpt-4o	zero_shot_role_play	1	1,0	3	0,4839	0,4839	0,0000	0,0000	10,03	1024
150	2	top50_rand	gpt-4o	zero_shot_chain_of_thought	1	1,0	3	1,0000	1,0000	0,0000	0,0000	26,72	1415
151	2	top50_rand	gpt-4o-mini	few_shot	1	1,0	3	0,5000	1,0000	1,0000	1,0000	7,56	1171

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
152	2	top50_rand	gpt-4o-mini	zero_shot	1	1,0	3	0,5000	1,0000	1,0000	1,0000	6,45	1050
153	2	top50_rand	gpt-4o-mini	few_shot_role_play	1	1,0	3	0,5000	1,0000	1,0000	1,0000	11,04	1178
154	2	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	1	1,0	3	0,5000	1,0000	1,0000	1,0000	8,45	1185
155	2	top50_rand	gpt-4o-mini	zero_shot_role_play	1	1,0	3	0,5000	1,0000	1,0000	1,0000	5,32	1060
156	2	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	1	1,0	3	0,5000	1,0000	1,0000	1,0000	11,22	1136
157	2	top50_rand	gpt-4o	few_shot	5	1,0	3	0,3372	0,5874	1,1832	1,0000	32,33	1327
158	2	top50_rand	gpt-4o	zero_shot	5	1,0	3	0,1747	0,3686	1,2649	1,2000	21,88	1079
159	2	top50_rand	gpt-4o	few_shot_role_play	5	1,0	3	0,2937	0,5874	1,0000	1,0000	12,83	1320
160	2	top50_rand	gpt-4o	few_shot_chain_of_thought	5	1,0	3	0,3191	0,5874	1,4142	1,2000	16,78	1366
161	2	top50_rand	gpt-4o	zero_shot_role_play	5	1,0	3	0,4621	0,5267	1,8439	1,4000	23,75	1273
162	2	top50_rand	gpt-4o	zero_shot_chain_of_thought	5	1,0	3	0,2863	0,5874	1,2649	1,2000	31,83	1705
163	2	top50_rand	gpt-4o-mini	few_shot	5	1,0	3	0,2937	0,5874	1,0000	1,0000	11,07	1433
164	2	top50_rand	gpt-4o-mini	zero_shot	5	1,0	3	0,4969	0,5874	1,0954	0,8000	7,95	1166
165	2	top50_rand	gpt-4o-mini	few_shot_role_play	5	1,0	3	0,3678	0,7355	1,0000	1,0000	10,36	1459
166	2	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	5	1,0	3	0,3273	0,5874	1,1832	1,0000	10,61	1363
167	2	top50_rand	gpt-4o-mini	zero_shot_role_play	5	1,0	3	0,2780	0,5874	1,4832	1,4000	11,22	1365
168	2	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	5	1,0	3	0,3455	0,5874	0,8944	0,8000	11,06	1357
169	2	top50_rand	gpt-4o	few_shot	10	1,0	3	0,1518	0,2833	1,8708	1,7000	22,24	1542
170	2	top50_rand	gpt-4o	zero_shot	10	1,0	3	0,1129	0,2648	2,1448	2,0000	14,25	1397
171	2	top50_rand	gpt-4o	few_shot_role_play	10	1,0	3	0,3622	0,4504	1,0488	0,7000	24,54	1713
172	2	top50_rand	gpt-4o	few_shot_chain_of_thought	10	1,0	3	0,6127	0,6924	0,8944	0,6000	20,88	1691
173	2	top50_rand	gpt-4o	zero_shot_role_play	10	1,0	3	0,3846	0,5069	1,3038	1,1000	21,18	1450
174	2	top50_rand	gpt-4o	zero_shot_chain_of_thought	10	1,0	3	0,4982	0,6356	0,8944	0,6000	24,60	1586
176	2	top50_rand	gpt-4o-mini	zero_shot	10	1,0	3	0,2503	0,4767	1,4491	1,3000	8,85	1192
177	2	top50_rand	gpt-4o-mini	few_shot_role_play	10	1,0	3	0,2403	0,4901	1,6125	1,4000	13,56	1337
178	2	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	10	1,0	3	0,2409	0,5069	1,3784	1,3000	8,97	1428
179	2	top50_rand	gpt-4o-mini	zero_shot_role_play	10	1,0	3	0,2920	0,4901	1,7321	1,4000	10,62	1279
180	2	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	10	1,0	3	0,2666	0,4987	1,2247	1,1000	10,58	1322
181	2	top50_rand	gpt-4o	few_shot	1	1,0	5	1,0000	1,0000	0,0000	0,0000	14,46	1240
183	2	top50_rand	gpt-4o	few_shot_role_play	1	1,0	5	0,5785	0,6854	0,7071	0,5000	14,66	1242
184	2	top50_rand	gpt-4o	few_shot_chain_of_thought	1	1,0	5	1,0000	1,0000	0,0000	0,0000	8,65	1197
185	2	top50_rand	gpt-4o	zero_shot_role_play	1	1,0	5	1,0000	1,0000	0,0000	0,0000	8,44	1063
186	2	top50_rand	gpt-4o	zero_shot_chain_of_thought	1	1,0	5	0,5000	1,0000	1,0000	1,0000	15,41	1491
187	2	top50_rand	gpt-4o-mini	few_shot	1	1,0	5	0,4839	0,4839	0,0000	0,0000	13,56	1629
188	2	top50_rand	gpt-4o-mini	zero_shot	1	1,0	5	0,2419	0,4839	1,0000	1,0000	6,81	1072
189	2	top50_rand	gpt-4o-mini	few_shot_role_play	1	1,0	5	0,2419	0,4839	1,0000	1,0000	8,45	1270

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
190	2	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	1	1,0	5	0,2419	0,4839	1,0000	1,0000	6,45	1204
191	2	top50_rand	gpt-4o-mini	zero_shot_role_play	1	1,0	5	0,2419	0,4839	1,0000	1,0000	7,58	1179
192	2	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	1	1,0	5	0,4839	0,4839	0,0000	0,0000	13,87	1472
193	2	top50_rand	gpt-4o	few_shot	5	1,0	5	0,4476	0,5874	1,4142	1,2000	14,16	1358
194	2	top50_rand	gpt-4o	zero_shot	5	1,0	5	0,4960	0,5874	1,0954	0,8000	24,99	1323
195	2	top50_rand	gpt-4o	few_shot_role_play	5	1,0	5	0,4476	0,5874	1,4142	1,2000	9,76	1273
196	2	top50_rand	gpt-4o	few_shot_chain_of_thought	5	1,0	5	0,5373	0,7355	0,8944	0,8000	15,62	1507
197	2	top50_rand	gpt-4o	zero_shot_role_play	5	1,0	5	0,4186	0,5267	1,7321	1,4000	27,44	1308
198	2	top50_rand	gpt-4o	zero_shot_chain_of_thought	5	1,0	5	0,2423	0,4123	1,4142	1,2000	14,01	1277
201	2	top50_rand	gpt-4o-mini	few_shot_role_play	5	1,0	5	0,1905	0,4123	1,4832	1,4000	10,86	1336
202	2	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	5	1,0	5	0,5018	0,5704	1,4832	1,0000	12,78	1334
203	2	top50_rand	gpt-4o-mini	zero_shot_role_play	5	1,0	5	0,4704	0,5267	1,6733	1,2000	9,69	1260
204	2	top50_rand	gpt-4o-mini	zero_shot_chain_of_thought	5	1,0	5	0,5150	0,5874	0,7746	0,6000	16,56	1440
205	2	top50_rand	gpt-4o	few_shot	10	1,0	5	0,2969	0,4238	1,6432	1,5000	13,12	1377
206	2	top50_rand	gpt-4o	zero_shot	10	1,0	5	0,5037	0,6177	1,0000	0,8000	19,12	1705
207	2	top50_rand	gpt-4o	few_shot_role_play	10	1,0	5	0,4035	0,5618	1,4832	1,2000	13,31	1539
209	2	top50_rand	gpt-4o	zero_shot_role_play	10	1,0	5	0,4245	0,5988	1,1832	1,0000	14,80	1592
210	2	top50_rand	gpt-4o	zero_shot_chain_of_thought	10	1,0	5	0,3389	0,4068	2,2804	1,8000	12,65	1405
211	2	top50_rand	gpt-4o-mini	few_shot	10	1,0	5	0,4053	0,5988	1,0954	1,0000	10,79	1395
212	2	top50_rand	gpt-4o-mini	zero_shot	10	1,0	5	0,3553	0,5906	1,3416	1,2000	7,29	1138
213	2	top50_rand	gpt-4o-mini	few_shot_role_play	10	1,0	5	0,4831	0,6474	1,1832	1,0000	8,68	1342
214	2	top50_rand	gpt-4o-mini	few_shot_chain_of_thought	10	1,0	5	0,4156	0,5906	1,4832	1,2000	8,13	1338
215	2	top50_rand	gpt-4o-mini	zero_shot_role_play	10	1,0	5	0,4678	0,5820	1,0801	0,8333	9,22	1282
217	2	most_common	gpt-4o	few_shot	1	1,0	3	0,5000	1,0000	1,0000	1,0000	10,63	1183
218	2	most_common	gpt-4o	zero_shot	1	1,0	3	0,5000	1,0000	1,0000	1,0000	6,93	948
219	2	most_common	gpt-4o	few_shot_role_play	1	1,0	3	0,5000	1,0000	1,0000	1,0000	13,38	1439
220	2	most_common	gpt-4o	few_shot_chain_of_thought	1	1,0	3	0,5000	1,0000	1,0000	1,0000	15,36	1152
221	2	most_common	gpt-4o	zero_shot_role_play	1	1,0	3	0,5000	1,0000	1,0000	1,0000	24,29	1104
222	2	most_common	gpt-4o	zero_shot_chain_of_thought	1	1,0	3	0,5000	1,0000	1,0000	1,0000	26,53	1079
223	2	most_common	gpt-4o-mini	few_shot	1	1,0	3	0,5000	1,0000	1,0000	1,0000	7,15	1099
224	2	most_common	gpt-4o-mini	zero_shot	1	1,0	3	0,5000	1,0000	1,0000	1,0000	98,03	994
225	2	most_common	gpt-4o-mini	few_shot_role_play	1	1,0	3	0,5000	1,0000	1,0000	1,0000	6,60	1169
226	2	most_common	gpt-4o-mini	few_shot_chain_of_thought	1	1,0	3	0,5000	1,0000	1,0000	1,0000	6,42	1169
227	2	most_common	gpt-4o-mini	zero_shot_role_play	1	1,0	3	0,5000	1,0000	1,0000	1,0000	6,09	1013
228	2	most_common	gpt-4o-mini	zero_shot_chain_of_thought	1	1,0	3	0,5000	1,0000	1,0000	1,0000	6,77	1050
229	2	most_common	gpt-4o	few_shot	5	1,0	3	0,6828	0,8569	0,7746	0,6000	15,67	1471

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
230	2	most_common	gpt-4o	zero_shot	5	1,0	3	0,5727	0,7355	0,7746	0,6000	11,30	1221
231	2	most_common	gpt-4o	few_shot_role_play	5	1,0	3	0,5487	0,5874	1,0000	0,6000	16,27	1504
232	2	most_common	gpt-4o	few_shot_chain_of_thought	5	1,0	3	0,5150	0,5874	0,7746	0,6000	13,54	1369
233	2	most_common	gpt-4o	zero_shot_role_play	5	1,0	3	0,7368	0,8569	0,6325	0,4000	15,88	1284
234	2	most_common	gpt-4o	zero_shot_chain_of_thought	5	1,0	3	0,7067	0,8003	0,7071	0,5000	17,46	1356
235	2	most_common	gpt-4o-mini	few_shot	5	1,0	3	0,2863	0,5874	1,2649	1,2000	8,60	1261
236	2	most_common	gpt-4o-mini	zero_shot	5	1,0	3	0,5158	0,7355	0,7746	0,6000	9,50	1251
237	2	most_common	gpt-4o-mini	few_shot_role_play	5	1,0	3	0,3265	0,5874	1,1832	1,0000	10,14	1379
238	2	most_common	gpt-4o-mini	few_shot_chain_of_thought	5	1,0	3	0,3265	0,5874	1,1832	1,0000	8,99	1297
239	2	most_common	gpt-4o-mini	zero_shot_role_play	5	1,0	3	0,3273	0,5874	1,1832	1,0000	10,95	1263
240	2	most_common	gpt-4o-mini	zero_shot_chain_of_thought	5	1,0	3	0,3479	0,6251	0,8944	0,8000	9,80	1186
241	2	most_common	gpt-4o	few_shot	10	1,0	3	0,3808	0,4604	1,5492	1,2000	14,61	1464
242	2	most_common	gpt-4o	zero_shot	10	1,0	3	0,7151	1,0000	2,3238	1,4000	18,83	1644
243	2	most_common	gpt-4o	few_shot_role_play	10	1,0	3	0,3710	0,4849	1,6432	1,3000	24,15	1686
244	2	most_common	gpt-4o	few_shot_chain_of_thought	10	1,0	3	0,4190	0,4849	1,3416	1,0000	16,14	1245
245	2	most_common	gpt-4o	zero_shot_role_play	10	1,0	3	0,2279	0,4253	2,0494	1,8000	15,26	1210
246	2	most_common	gpt-4o	zero_shot_chain_of_thought	10	1,0	3	0,4574	0,6177	1,2247	0,9000	20,13	1252
247	2	most_common	gpt-4o-mini	few_shot	10	1,0	3	0,4073	0,6924	1,9494	1,4000	8,97	1342
248	2	most_common	gpt-4o-mini	zero_shot	10	1,0	3	0,2842	0,5906	2,0736	1,7000	8,52	1218
249	2	most_common	gpt-4o-mini	few_shot_role_play	10	1,0	3	0,3219	0,5906	1,5811	1,3000	23,36	1366
250	2	most_common	gpt-4o-mini	few_shot_chain_of_thought	10	1,0	3	0,2875	0,5618	1,7607	1,5000	12,56	1312
251	2	most_common	gpt-4o-mini	zero_shot_role_play	10	1,0	3	0,3122	0,5906	1,4142	1,2000	12,55	1273
252	2	most_common	gpt-4o-mini	zero_shot_chain_of_thought	10	1,0	3	0,4033	0,6745	1,8974	1,4000	10,24	1226
253	2	most_common	gpt-4o	few_shot	1	1,0	5	1,0000	1,0000	0,0000	0,0000	12,45	1428
254	2	most_common	gpt-4o	zero_shot	1	1,0	5	0,5000	1,0000	1,0000	1,0000	16,27	1330
255	2	most_common	gpt-4o	few_shot_role_play	1	1,0	5	1,0000	1,0000	0,0000	0,0000	7,10	1212
256	2	most_common	gpt-4o	few_shot_chain_of_thought	1	1,0	5	0,5000	1,0000	1,0000	1,0000	9,77	1385
257	2	most_common	gpt-4o	zero_shot_role_play	1	1,0	5	1,0000	1,0000	0,0000	0,0000	20,20	1483
258	2	most_common	gpt-4o	zero_shot_chain_of_thought	1	1,0	5	1,0000	1,0000	0,0000	0,0000	18,47	1317
259	2	most_common	gpt-4o-mini	few_shot	1	1,0	5	0,5000	1,0000	1,0000	1,0000	8,77	1141
260	2	most_common	gpt-4o-mini	zero_shot	1	1,0	5	0,5000	1,0000	1,0000	1,0000	5,71	1022
261	2	most_common	gpt-4o-mini	few_shot_role_play	1	1,0	5	0,5000	1,0000	1,0000	1,0000	9,47	1161
262	2	most_common	gpt-4o-mini	few_shot_chain_of_thought	1	1,0	5	0,5000	1,0000	1,0000	1,0000	7,77	1201
263	2	most_common	gpt-4o-mini	zero_shot_role_play	1	1,0	5	0,5000	1,0000	1,0000	1,0000	6,57	1065
264	2	most_common	gpt-4o-mini	zero_shot_chain_of_thought	1	1,0	5	0,5000	1,0000	1,0000	1,0000	8,22	1104
265	2	most_common	gpt-4o	few_shot	5	1,0	5	0,5675	0,7694	1,8974	1,2000	11,21	1188

EK 10 Deneysel sonuçlar (devam)

scenario	run	ingredient_selection	gpt_version	prompt_type	num_recipes	gpt_temperature	top_n	WNDCG	NDCG	rmse	mae	response_time	total_tokens
266	2	most_common	gpt-4o	zero_shot	5	1,0	5	0,4271	0,6978	1,4142	1,2000	11,87	1224
267	2	most_common	gpt-4o	few_shot_role_play	5	1,0	5	0,5404	0,5874	1,2649	0,8000	14,63	1282
268	2	most_common	gpt-4o	few_shot_chain_of_thought	5	1,0	5	0,5487	0,5874	1,0000	0,6000	11,59	1263
269	2	most_common	gpt-4o	zero_shot_role_play	5	1,0	5	0,6544	0,7355	1,8439	1,0000	10,90	1210
270	2	most_common	gpt-4o	zero_shot_chain_of_thought	5	1,0	5	0,5478	0,5874	1,0000	0,6000	10,87	1253
271	2	most_common	gpt-4o-mini	few_shot	5	1,0	5	0,2854	0,5874	1,2649	1,2000	9,26	1286
272	2	most_common	gpt-4o-mini	zero_shot	5	1,0	5	0,3462	0,6251	1,1832	1,0000	9,07	1217
273	2	most_common	gpt-4o-mini	few_shot_role_play	5	1,0	5	0,3462	0,6251	1,1832	1,0000	10,03	1377
274	2	most_common	gpt-4o-mini	few_shot_chain_of_thought	5	1,0	5	0,3462	0,6251	1,1832	1,0000	9,07	1341
275	2	most_common	gpt-4o-mini	zero_shot_role_play	5	1,0	5	0,5027	0,7355	1,0954	0,8000	8,36	1156
276	2	most_common	gpt-4o-mini	zero_shot_chain_of_thought	5	1,0	5	0,2780	0,5874	1,4832	1,4000	8,90	1317
277	2	most_common	gpt-4o	few_shot	10	1,0	5	0,4061	0,5785	1,6432	1,3000	13,04	1312
278	2	most_common	gpt-4o	zero_shot	10	1,0	5	0,4308	0,5785	1,5492	1,2000	18,44	1352
279	2	most_common	gpt-4o	few_shot_role_play	10	1,0	5	0,3710	0,5566	1,6733	1,4000	20,60	1633
280	2	most_common	gpt-4o	few_shot_chain_of_thought	10	1,0	5	0,4058	0,4987	1,2649	1,0000	16,54	1341
281	2	most_common	gpt-4o	zero_shot_role_play	10	1,0	5	0,3834	0,5069	1,7607	1,3000	22,45	1475
282	2	most_common	gpt-4o	zero_shot_chain_of_thought	10	1,0	5	0,5991	0,7749	0,7746	0,6000	22,71	1483
284	2	most_common	gpt-4o-mini	zero_shot	10	1,0	5	0,3839	0,5820	1,1832	1,0000	10,67	1251
285	2	most_common	gpt-4o-mini	few_shot_role_play	10	1,0	5	0,3951	0,5820	1,2649	1,0000	8,36	1260
286	2	most_common	gpt-4o-mini	few_shot_chain_of_thought	10	1,0	5	0,3999	0,6388	1,4832	1,2000	8,51	1336
287	2	most_common	gpt-4o-mini	zero_shot_role_play	10	1,0	5	0,2815	0,5906	1,8439	1,6000	6,57	1074
288	2	most_common	gpt-4o-mini	zero_shot_chain_of_thought	10	1,0	5	0,4085	0,6177	1,3416	1,0000	8,76	1247