

TOBB Ekonomi ve Teknoloji Üniversitesi Bilgisayar Mühendisliği Bölümü

BİL 496 – BİTİRME PROJESİ

TRAINER

Beyza Altanlar - 151101040

İÇİNDEKİLER

Projenin amacı	2
Hedef Kitle	2
Gereksinim Analizi	3
Kullanıcı Senaryoları	4
Başarılı Sonuç İçin Kriterler	5
Sistem Fonksiyonları	6
Örnek Girdi Çıktı	8
Program Açıklaması	12
Çalışma Mantığı	12
Sistemin Detaylı Açıklaması	13
Machine Learning	13
Sonuç ve Değerlendirme(Machine Learning)	16
Program Oluşturma için Fine Tuning Kullanımı v	e Yorumu.18
ChatBot:Fine Tuning	20
Geliştirme	20
Harcamalar-Linkler	21

PROJENÍN AMACI

Trainer, kullanıcıların sağlık durumlarını sakatlıklarını göz önünde bulundurarak riske girmeden, yaşlarına, kilolarına ve hedeflerine uygun olarak kişiselleştirilmiş egzersiz programları oluşturmak ve sunmaktır. Her kullanıcının benzersiz ihtiyaçları ve hedefleri göz önüne alınarak tasarlanan bu antrenman programları, sağlıklı bir yaşam tarzını teşvik etmek ve kullanıcıların hedeflerine ulaşmalarına yardımcı olmak için tasarlanmıştır.

HEDEF KITLE

Proje, geniş bir kullanıcı kitlesine hitap etmeyi hedeflemektedir. Yaşları, cinsiyetleri, sağlık durumları ve yaşam tarzları farklı olan herkes, bu uygulamayı kullanarak kendi ihtiyaçlarına uygun egzersiz programlarına erişebileceklerdir.

Egzersiz yaparken dikkat edilmesi gereken özel sağlık durumlarına sahip bireyler de projemizin hedef kitlesini oluşturmaktadır. Egzersizler kafa/boyun bölgesi, göğüs bölgesi, bacak bölgesi, ayak bölgesi ve kol bölgesi olarak buralarda sakatlık durumlarına uygun olarak önerilmektedir.

Egzersiz yaparken danışmak isteyen ve beslenmesine de dikkat etmek isteyen kişilere destek için bir chatbot da oluşturduk.

GEREKSİNİM ANALİZİ

Proje Gereksinim Analizi: Kişiselleştirilmiş Egzersiz ve Beslenme Uygulaması

Kullanıcı Girişi ve Profil Oluşturma:

Kullanıcıların uygulamaya kaydolabilmesi ve giriş yapabilmesi için bir kullanıcı yönetim sistemi gereklidir. Kullanıcıların profil oluşturması için temel bilgilerin (kilo, boy, sağlık durumu vb.) toplanması gerekmektedir.

Sağlık Durumu ve Hedef Belirleme:

Kullanıcıların sağlık durumlarını değerlendirmek ve hedeflerini belirlemek için bir anket veya form sunulmalıdır. Sağlık durumu ve hedeflere göre kişiye özel egzersiz ve beslenme programları oluşturulmalıdır.

Kişiselleştirilmiş Egzersiz Programları:

Kullanıcıların sağlık durumlarına, yaşlarına, hedeflerine ve egzersiz geçmişlerine göre kişiselleştirilmiş egzersiz programları oluşturulmalıdır. Egzersiz programları, farklı egzersiz türlerini ve seviyelerini içermelidir.

Beslenme Danışmanlığı:

Kullanıcıların sağlık durumları ve hedeflerine göre kişiselleştirilmiş beslenme önerileri sunulmalıdır. Beslenme danışmanlığı, kullanıcıların günlük kalori ihtiyaçlarını, besin gruplarını ve önerilen öğün planlarını içermelidir.

Chatbot ve Yardım Sistemi:

Kullanıcıların egzersiz ve beslenme ile ilgili sorularını cevaplamak için bir chatbot veya yardım sistemi entegre edilmelidir. Chatbot, kullanıcıların sorularını anlayabilecek ve doğru yanıtlar verebilecek şekilde eğitilmelidir. Chatgpt kullanılarak finetuning ile özelleştirilmiştir.

Veri Analizi ve Geri Bildirim:

Kullanıcıların egzersiz ve beslenme alışkanlıklarını takip etmek için bir veri analizi sistemi gereklidir. Veri analizi, kullanıcıların ilerlemelerini izlemek, önerileri güncellemek ve daha iyi sonuçlar elde etmek için kullanılmalıdır.

KULLANICI SENARYOLARI

Kullanıcı, sağlıklı yaşamaya yönelik bir uygulama olan "Trainer" uygulamasını indirdi ve kayıt işlemlerini tamamladıktan sonra giriş yaptı. Ana ekranda, "Egzersiz Programı Seçimi" adlı bir seçenek bulunmaktadır. Kullanıcı, sağlık durumu ve hedeflerine uygun bir egzersiz programı seçmek için bu seçeneği tıklar. Uygulama, kullanıcıya bir dizi soru sorarak egzersiz programını kişiselleştirmesine yardımcı olur. Kullanıcıya kilosunu, boyunu ve sağlık durumunu belirten cesitli sorular yöneltilir. Kullanıcı, sağlık durumu ve hedeflerine bağlı olarak egzersiz programı seçimini tamamlar. Örneğin, kullanıcı kilo vermek istiyorsa ve belirli bir sağlık sorunuyla mücadele ediyorsa, uygulama bu bilgilere dayanarak uygun bir program önerir. Trainer, kullanıcının verdiği bilgilere göre kişiselleştirilmiş bir egzersiz programı oluşturur. Egzersiz programı, kullanıcının yaşına, sağlık durumuna ve hedeflerine uygun olarak kardiyo, ağırlık antrenmanı, pilates, esneme gibi farklı egzersiz türlerini içerebilir. Kullanıcı, önerilen egzersiz programını inceleyebilir ve kabul ettiği takdirde uygulama üzerinden programı görebilir. Egzersiz programı, kullanıcının uygunluğuna göre planlanmış olabilir. Kullanıcı, Trainer uygulaması aracılığıyla egzersiz yapmaya başlar ve programı düzenli olarak takip eder.

Başarılı Sonuç için Kriterler

- Uygulamadan beklediğimiz program, sakatlığı olan birine; örneğin dizde problemi olan birine deadlift gibi bacak hareketleri önermemelidir.
 chatbotta çölyak birine glutenli diyet önermemelidir, kilo vermek ve yağ yakmak isteyen kişiye yağlı ürünler önermemelidir. Kişiselleştirme ve uyarlanabilirlik önemlidir.
- Geri Bildirim ve İletişim: Kullanıcıların uygulama ile etkileşimde bulunmaları ve geri bildirimde bulunmaları teşvik edilmelidir. Kullanıcıların soruları ve geri bildirimleri hızlı ve etkili bir şekilde yanıtlanmalıdır. chatbot kısmında kullanıcı geri bildirimde bulunup yeni hareket önerisi isteyebilir.
- Etkinlik ve İyileşme: Kullanıcıların sağlık ve performansındaki iyileşmeleri ölçmek için uygun metrikler kullanılmalıdır. Egzersiz ve beslenme programlarının etkinliği, kullanıcıların hedeflerine ulaşma ve sağlık durumlarında olumlu değişiklikler yapma yeteneği ile ölçülebilirdi. Bununla ilgili bir şey geliştiremedik. Programda "Kullanıcı takibi" yapılabilirdi.
- Doğruluk ve Güvenilirlik: Egzersiz ve beslenme önerileri, bilimsel kanıtlara dayanmalı ve güvenilir kaynaklardan alınmalıdır. Veri analizi ve öneriler, doğru ve güncel bilgilere dayanmalı. Bunun dışında karmaşık bir yapı kullanılabilirdi, model doğruluğumuzun oranı artırılabilir. Projemiz için veri satın almadık. Daha büyük ve temiz bir veri kümesi kullanıldığında model doğruluğu artacaktır. Bunun dışında daha büyük bir veriyle ensemble modeller denebilirdi.

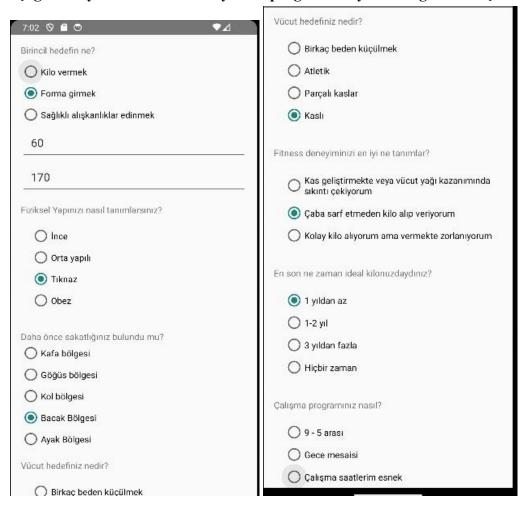
SİSTEM FONKSİYONLARI

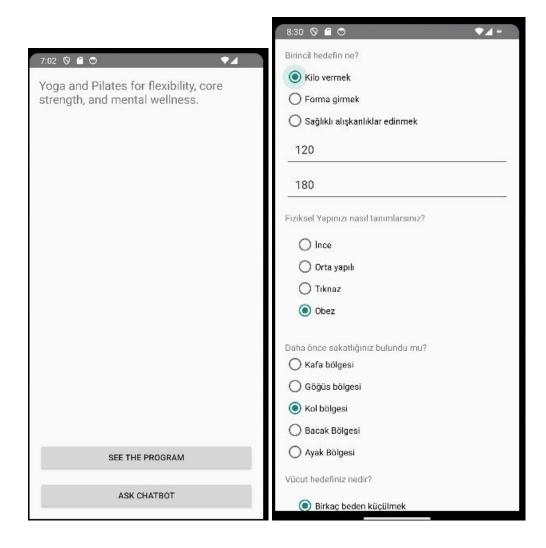


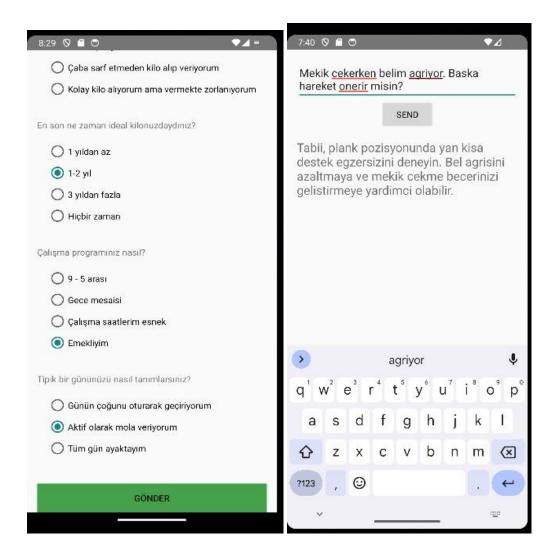
En son ne zaman ideal kilonuzdaydınız?	Birincil hedefin ne?
1 yıldan az	→ Kilo vermek → Forma girmek
3 yıldan fazla	Sağlıklı alışkanlıklar edinmek
O Hiçbir zaman	Kilonuzu girin
Çalışma programınız nasıl?	Boyunuzu girin (cm)
9 - 5 arası Gece mesaisi	Fiziksel Yapınızı nasıl tanımlarsınız?
Çalışma saatlerim esnek	ince Orta yapılı
○ Emekliyim	Tiknaz
Tipik bir gününüzü nasıl tanımlarsınız?	Obez
Günün çoğunu oturarak geçiriyorum	Daha önce sakatlığınız bulundu mu?
Aktif olarak mola veriyorum	Kafa bölgesi
Tüm gün ayaktayım	Göğüs bölgesi
	Col bölgesi
GÖNDER	Bacak Bölgesi

ÖRNEK GİRDİ ÇIKTI

Aşağıda 2 ayrı örnek kombinasyonlar program arayüzünde gösterilmiştir:









PROGRAM AÇIKLAMASI

1. ÇALIŞMA MANTIĞI:

Uygulamaya giriş yaptığımızda bizi yukarıdaki ilk ekranlar karşılıyor. Sisteme kayıt olduktan sonra, kullanıcı maili ve şifresiyle giriş yapıyor. Daha sonra kullanıcının sakatlık (sağlık) durumunu, hedeflerini, aktiflik oranını (sedanter yaşam veya hareketli yaşam) ve kendisine uygun program sunabilmek için yoğunluk düzeyini alıyoruz.

Kullanıcıdan girdileri aldıktan sonra kısıtlara uygun bir mantık çerçevesinde bir spor ataması yapar. (kardiyo , ağırlık antrenmanı, pilates, esneme vb) Örneğin vücut ağırlığıyla antrenman. Daha sonra kullanıcı, programı görebilir veya direkt chatbot ile konuşabilir. Program çıktısı kullanıcının fiziksel anlamda sağlık sorununa göre ve yaşam tarzına göre Machine Learning kullanarak tasarlanmıştır.

Oluşturduğumuz chatbota yaptığı antrenmanla ilgili danışabileceği gibi, beslenme diyetiyle ilgili de danışabilir. Bir diyet/spor asistanı olarak oluşturduğumuz chatbot, yine kullanıcının kısıtlarına bağlı olarak çıktı oluşturur. Kişi şeker hastası veya çölyak gibi sağlık sorunlarına sahip olabilir. Çıktı buna yönelik öneriler yapar.

2. SİSTEMİN DETAYLI AÇIKLAMASI

MACHINE LEARNING

Projemizde kullanılan Machine Learning algoritması, kullanıcıların sağlık durumlarına, yaşlarına, fiziksel yeteneklerine ve hedeflerine göre kişiselleştirilmiş egzersiz programları oluşturur. Verilerin yüzde 20sini test, yüzde 80 i eğitmek için kullanılmıştır. Sınıflandırma problemi olduğu için accuracy değerleri üzerinden karşılaştırma yapıldı. Kullanılan metrikler; accuracy, precision, Cross validation'dır. Veriler arası corelasyon ölçümü yapıldı.

Logistic Regression: Modelin doğruluğu (accuracy) ve hassasiyeti (precision) değerlendirildiğinde, test seti üzerinde %30 doğruluk ve %30 hassasiyet elde edildi. Decision Tree: Decision Tree modeli, test seti üzerinde %31.5 doğruluk ve %30.4 hassasiyet elde etti. Model parametrelerine dikkat ederek, farklı modeller deneyerek veya ensemble modeller oluşturarak doğruluğu artırmak amaçlandı. Random Forest Model tek başına kullanıldığında doğruluk oranı 0.36 olarak görüldü. Ek olarak, sayısal değişkenler arasındaki korelasyonu incelemek için bir korelasyon matrisi gösterildi. Korelasyon matrisine baktığımızda, "weight" (kilo) ve "height" (boy) değişkenleri arasında çok düşük bir korelasyon olduğunu görüyoruz. Bu, bu iki değişken arasında güçlü bir ilişki olmadığını ve birbirlerini belirgin şekilde etkilemediğini gösterir.

weight height weight 1.000000 0.028543 height 0.028543 1.000000

Verideki özellikler(features) baştan gözden geçirildi, bütün veriler kontrol edilip tutarsızlıklar düzeltildi. Uygun olmayan değerler değiştirildi. Tekrar sonuçlar alındı. Logistic Regression %35 doğruluk oranı verdi.

Gradient ve XGBoost Model doğruluk oranları % 36 olarak görüldü. SVM model ile denendiğinde accuracy 0.195 alındı. Veri temizlendikten sonra sonucun SVM modeli için değişmediği görüldü. Daha sonra, SVM modeli RBF (Gaussian) çekirdeği kullanarak oluşturup ve eğitildi. C ve gamma parametreleri varsayılan değerlerle ayarlandı, SVM modelinde sonuç diğerlerinden açıkça bir farkla düşük kaldı.

Neural Network Model Accuracy: 0.385 ile diğer modellere görece yüksek bir oran yakalandı. Gaussian Naive Bayes algoritması kullanılarak alınan doğruluk oranı 0.33 oldu.

Random Forest modeli kullanıldığında 0.36 alınmıştı , veri düzenlenmesinden sonra sonuçlar

aşağıdaki gibi oldu : Random Forest Model Accuracy: 0.385

Cross-Validation Scores: [0.2875 0.2625 0.25 0.29375 0.28125]

Hassasiyet: 0.3450506864623244

Hiperparametre optimizasyonu için hiperparametre tuning yapıldı; bunun için GridSearchCV kullanıldı. GridSearchCV, belirli bir hiperparametre uzayını (örneğin, farklı ağaç derinlikleri veya yaprak düğümü sayıları gibi) ve bu hiperparametrelerin her bir kombinasyonu için çapraz doğrulama kullanarak modelin performansını değerlendirir. Bu şekilde en iyi hiperparametre kombinasyonu belirlendi. Bu durumda, en iyi hiperparametreler max_depth=20, min_samples_leaf=1, min_samples_split=2 ve n_estimators=100 olarak bulunmuştur. Random forest hiperparametre ayarlaması sonucu accuracy: 0.35122450320557946

Model doğruluk oranını artırmak için, birden çok modelin bir araya gelerek daha güçlü bir tahmin yapmasını sağlayan bir makine öğrenimi tekniği olan "Ensemble Modeller" kullandım. Hem iyi performans hem tahmin gücü açısından birleştirilmiş modeller denendi. daha genelleştirilmiş bir tahmin yapması beklendi. Ensemble Modellerde aldığım sonuçlar: Ensemble model (random forest, gradient): 0.37

Ensemble model (random forest, gradient, gaussian): 0.375

Ensemble model (decision, random, svm): 0.31

Ensemble model (gradient, random, svm): 0.325

Bu modeller Voting Classifier yöntemiyle bir araya getirip oluşturulmuştur. Modellerin tahminlerinin birleştirilip daha genelleştirilmiş bir tahmin yapmak amaçlandı.

Ensemble Learning tekniklerinden "Stacking" yöntemini deneyerek en iyi doğruluk oranını almayı başardım. Stacking özelinde, farklı temel (base) modelleri bir araya getirerek bir meta model oluşturulur. iki farklı temel model olarak: Logistic Regression (LogisticRegression(max_iter=1000)) ve Decision Tree (DecisionTreeClassifier()) kullanıldı. Doğruluk oranı % 40 olarak görüldü. meta model olarak RandomForestClassifier kullanılmıştır. Farklı türdeki modellerin farklı özelliklerini birleştirerek daha güçlü ve genelleştirilmiş bir model elde edilmiştir. Stacking ile ilgili başka model kombinasyonları denendiğinde oranlar: Random Forest ve Gradient birlikte kullanıldığında (Stacking yöntemiyle): 0.295 Gradient Boosting ve SVM birlikte kullanıldığında (Stacking Model): 0.38 olarak görülmüştür.

Diğer denenen modeller:

ANN: 0.33 KNN: 0.135

Başarımın artmasını veri düzenlemek, birkaç modeli bir arada kullanmak ve stacking yöntemini kullanmak sağlamıştır.

Sonuç Ve Değerlendirme (Machine Learning)

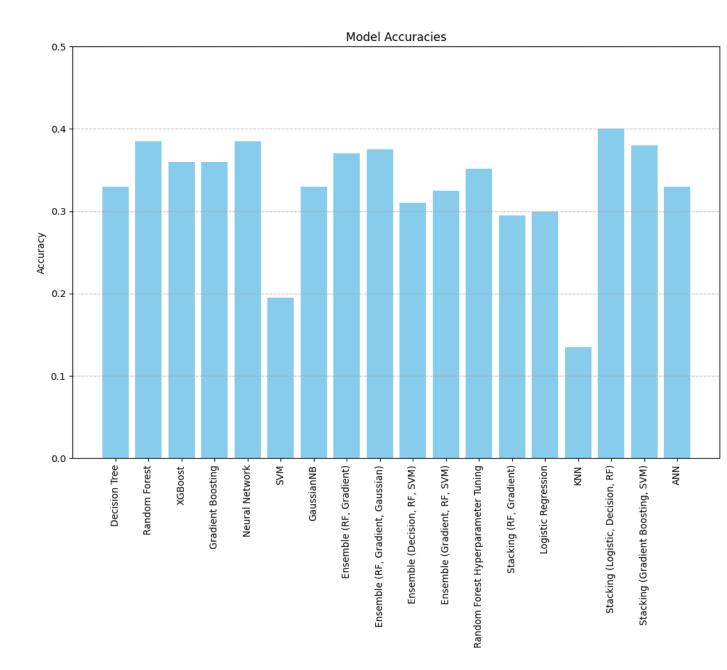
Sonuç olarak, bu çalışma kapsamında, sağlık durumu, yaş, fiziksel yetenekler ve hedefler gibi faktörlere dayalı olarak kişiselleştirilmiş egzersiz programlarının oluşturulması için çeşitli makine öğrenimi modelleri ve teknikleri değerlendirildi. İlk olarak, veri ön işleme adımları ile veri setindeki eksik veya tutarsız verilerin düzeltilmesi ve uygun olmayan değerlerin değiştirilmesi yapıldı. Daha sonra, farklı makine öğrenimi modelleri (Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, SVM, Neural Network, Gaussian Naive Bayes, XGBoost, Catboost) kullanılarak egzersiz programı önerileri için tahminler yapıldı ve bunlarla birlikte ensemble modeller kullanıldı.

Modellerin performansını değerlendirirken, doğruluk oranları, hassasiyet ve Cross-Validation skorları gibi performans metrikleri kullanıldı. Ayrıca, hiperparametre ayarlama yöntemleri ile en iyi model yapılandırmaları belirlendi. Bununla birlikte, ensemble modelleme teknikleri (Random Forest, Gradient Boosting, SVM) ve stacking gibi yöntemler kullanılarak model performansı artırılmaya çalışıldı.

Sonuçlar, ensemble modellerin ve stacking yönteminin, tek başına kullanılan modellere kıyasla daha iyi performans sağladığını göstermektedir. Özellikle, stacking yöntemi ile farklı temel modellerin bir araya getirilmesi, daha genelleştirilmiş ve güçlü tahminler yapılmasını sağlamıştır.

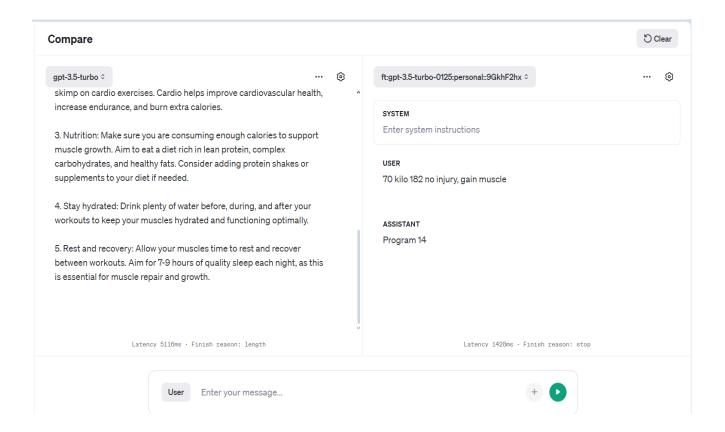
Ancak, elde edilen sonuçların gerçek dünya uygulamalarına genelleştirilmesi için daha fazla deneysel çalışma ve doğrulama gerekmektedir. Ayrıca, daha fazla veri toplanması ve modelin gerçek kullanıcılar üzerinde test edilmesi, modelin güvenilirliğini ve etkinliğini artırabilir.

Aşağıdaki tabloda accuracy değerleri görülmektedir.



Program Oluşturma için Fine Tuning Kullanımı ve Yorumu

Machine Learning Modellerine 2. bir alternatif olarak Fine tuning kullanıldı. Chatgpt de kullanılıp karşılaştırma yapıldı.



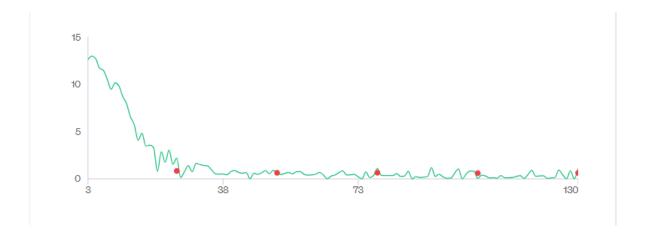
Training loss: 1.0595

Validation loss:

0.0826

Full validation loss:

0.6253



Eğitim kaybının, doğrulama kaybından daha yüksek olduğunu ve tam doğrulama kaybının doğrulama kaybından daha yüksek olduğunu görebiliriz.

Training Loss, eğitim sırasında modelin eğitim veri kümesine ne kadar iyi uyum sağladığını gösterir. Eğitim kaybı, modelin eğitim verilerini ne kadar iyi tahmin ettiğini ve ne kadar hızlı öğrendiğini yansıtır. Burada 1.0595 olarak gösterilen değer, modelin eğitim verilerine kıyasla tahminlerinde ortalama yaklaşık 1.06 birimlik bir hata yaptığını gösterir. Validation Loss, Modelin doğrulama veri kümesine ne kadar iyi uyum sağladığını gösterir. Doğrulama kaybı, modelin genelleme yeteneğini ve aşırı uyum (overfitting) durumunu değerlendirir. Burada 0.0826 olarak gösterilen değer, modelin doğrulama verilerini ne kadar iyi tahmin ettiğini ve genellenebilirlik düzeyini yansıtır. Düşük bir doğrulama kaybı, modelin iyi bir genelleme yapabildiğini gösterir. Full Validation Loss, Birden fazla doğrulama seti kullanılarak hesaplanan ve modelin genel doğrulama performansını ölçen bir değerdir. Burada 0.6253 olarak gösterilen değer, modelin tüm doğrulama veri kümelerine ortalama olarak yaklaşık 0.63 birimlik bir hata yaptığını gösterir.

Ancak, tam doğrulama kaybının doğrulama kaybından biraz daha yüksek olması, modelin tüm doğrulama veri kümelerine tam olarak genelleme yapamayabileceğini gösterebilir. Bu da overfittinge neden olabilir. Geniş bir veri kümesi kullanıldığında oranlar değişecektir.

ChatBot: Fine-Tuning

Chatbot için fine-tuning kullanarak, önceden eğitilmiş bir dil modeli üzerinde spesifik olarak bu projeye özelleştirerek chatgpt-3.5 kullandık. Chatbotun amacı, projemizin amacına uygun olarak kullanıcıların egzersiz ve beslenme ile ilgili sorularına cevap vermek ve kişiselleştirilmiş öneriler sunmaktı. Programımızı önerdikten sonra doğru ve tutarlı yanıtlar almak için uygun verilerle eğitmemiz gerekiyordu. Egzersiz ve beslenme alanında daha spesifik bir yetenek kazandırmak için projemizin diet.json dosyasındaki veri setiyle eğitildi. Chatgpt tarafından belli bir aralıkta veri seti kullanmamıza izin verildi. Bu haliyle soru cevaplarda başarılı sonuçlar elde ettik. Testlerimizde kısıtlara uymayan öneriler sunmadı.

Geliştirme:

Projemizde Java, Kotlin ve Python gibi programlama dillerini kullandık. Android uygulamasını Java ve Kotlin kullanarak geliştirdik, backend işlemlerinde Python, Java kullandık. Android uygulamasını geliştirmek için Android Studio geliştirme ortamı (IDE) kullandık. Backend işlemlerini yönetmek için ise Python'ı tercih ettik ve bu işlemleri Heroku gibi bulut tabanlı bir platformda sakladık. Heroku, Python ve diğer dillerle yazılmış uygulamaları kolayca dağıtmak ve yönetmek için kullanılan bir Platform Servis Sağlayıcısı (PaaS) olarak hizmet vermekte. Bu teknolojileri kullanarak, çok platformlu bir uygulama geliştirdik. Kullanıcı verilerini tutmak için FireBase veritabanını kullandık. Firebase seçmemizin nedeni kolay ve hızlı entegre olması ve Google tarafından sağlanmış bir database olmasıydı. Programları numaralandırdık ve Java kodumuzun içinde firebase'den programları getiriyoruz. ResultActivity Class'ında programı firebase'den çekip chatbotu da burada çağırıyoruz. Chatbot için OpenAI daki yönlendirmeleri kullandık.

Proje İçin Yapılan Harcamalar

Heroku 5\$ Chatgpt - Fine Tuning 5\$

VIDEO LINKI

https://www.youtube.com/watch?v=ERHngnaiRBA