

Kredi Temerrüt Tahmini ve Faiz Oranı Hesaplama Projesi

HT3

Zeynep Rana Başbüyük	21120205033
Ayşe Emine Erolur	21120205018
Ayşegül Sav	21120205027

Proje Amacı

Bu proje, kullanıcıdan alınan verilere dayanarak kredi temerrüt riskini öngören ve bu risk doğrultusunda kişiye özel faiz oranları hesaplayan bir sistem geliştirmeyi hedeflemektedir. Kredi sağlayıcı kurumların risk yönetim süreçlerini iyileştirmek ve müşterilere daha adil kredi teklifleri sunmalarını sağlamak temel amaçlar arasındadır. Bu bağlamda, mevcut literatürde yer alan makine öğrenimi tabanlı yöntemler, kredi başvurularını değerlendirme sürecinde yüksek doğruluk ve hızla tahminleme yapabilen etkili araçlar olarak öne çıkmaktadır[1]. Bu proje, finansal analiz işlevlerini ve kredi temerrüt riskini doğru bir şekilde öngörme yeteneğini birleştirerek, kullanıcıların kredi başvurularının onaylanma olasılıklarını daha doğru bir şekilde tahmin edebilmekte ve bu sayede kişiye özel faiz oranları hesaplamaktadır. Hem tahminleme hem de finansal analiz işlevlerini birleştiren bu proje, yenilikçi bir çözüm sunmayı planlamaktadır.

Proje Adımları

1. Standart Veri Bilimi İş Akışı

Problem tanımı

Günümüzde finansal kuruluşlar, kredi başvurularını değerlendirirken doğru kararlar alabilmek için büyük veri analizine dayalı yöntemlere ihtiyaç duymaktadır. Geleneksel kredi değerlendirme yöntemleri, genellikle zaman alıcıdır ve insan hatasına açıktır. Yanlış bir kredi tahmini, finansal kuruluşlar için hem maddi kayıplara hem de müşteri memnuniyetsizliğine yol açabilir. Bu projede, bir kredi başvurusunun onaylanıp onaylanmayacağını tahmin etmek için makine öğrenimi tekniklerinden yararlanarak daha hızlı, doğru ve etkili bir değerlendirme süreci yaratılışı için bir adım atıldı.

Veri Toplama

Bu proje kapsamında kullanılan veri seti, Credit-Risk Dataset [2] olarak adlandırılan, açık kaynaklı ve ikincil türden bir veri setidir. Birincil kaynaklı verilere erişim mümkün olmadığından, analizler için OpenML platformunda yayınlanmış olan bu veri setinden faydalanılmıştır. Veri setinde 32.581 örnek ve 12 özellik bulunmaktadır. Bu özellikler aşağıdaki

tabloda gösterilmiştir.

Tablo 1. Veri Seti Özellikleri

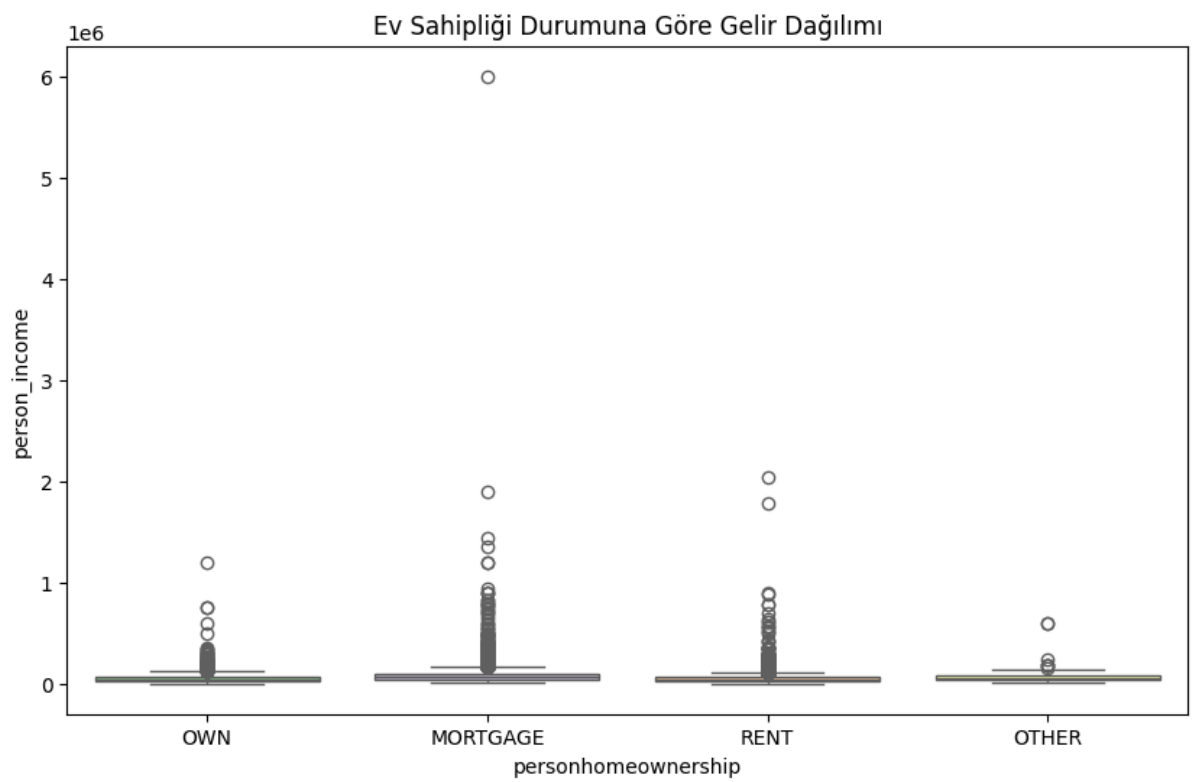
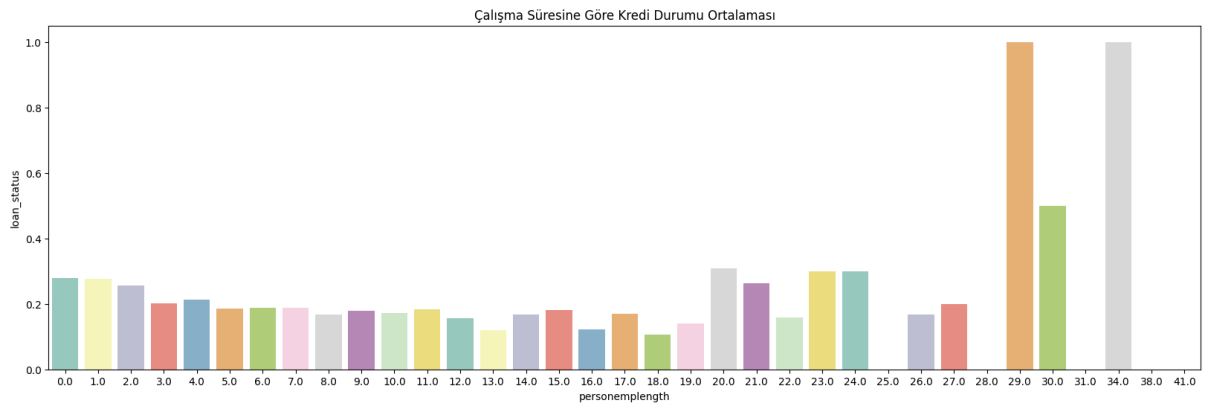
person_age	Kişinin yaşı	person_homeownership	Kişinin ev sahipliği durumu	loan_intent	Kreditin amacı
person_income	Kişinin geliri	person_employment_length	Kişinin işte çalışma süresi (yıl cinsinden)	loan_grade	Kredi derecesi
loan_amnt	Kredi tutarı	loan_int_rate	Kredi faiz oranı	loan_status	Kreditin durumu (ödenip ödenmediği)
loan_percent_income	Gelire oranla kredi yüzdesi	cb_person_default_on_file	Kredi geçmişinde temerrüt durumu	cb_person_credit_history	Kredi geçmişi uzunluğu (yıl cinsinden)

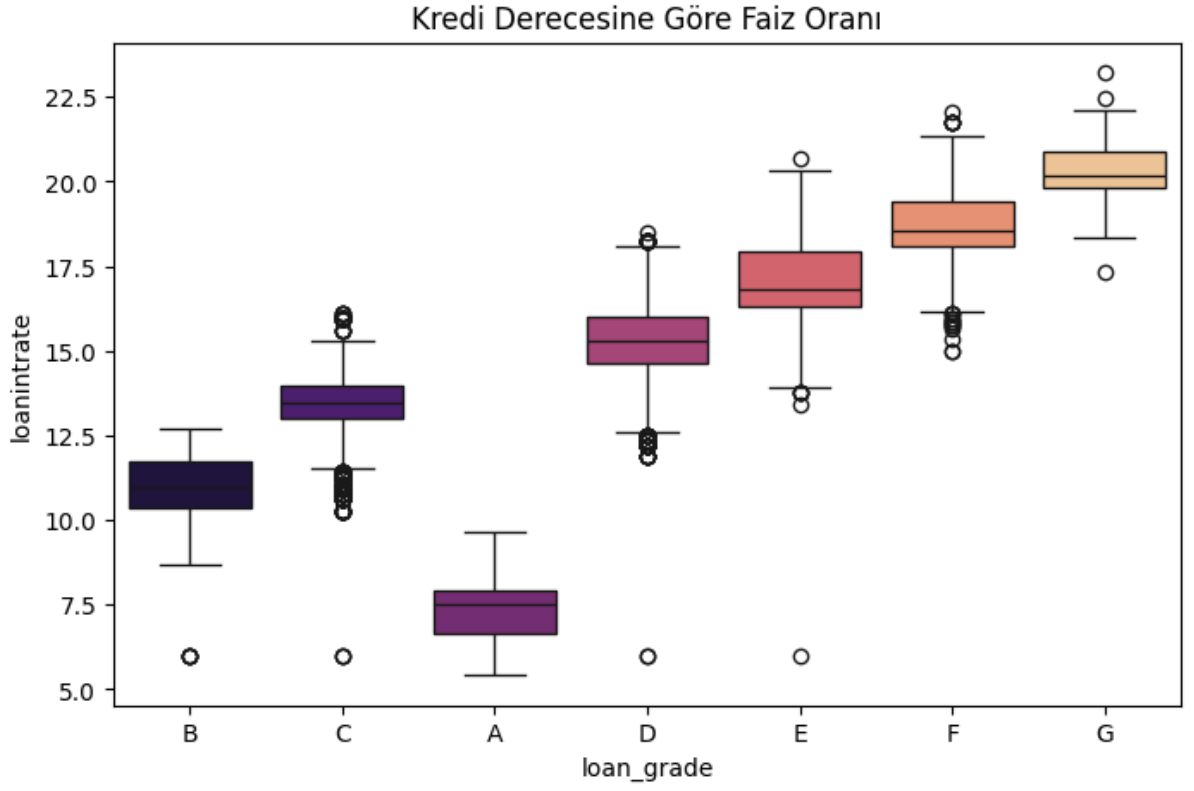
Veri Ön İşleme

- Veri kümesindeki boş veya anlamsız değerler tespit edilerek temizlenmiştir. Özellikle loanintrate (faiz oranı) ve personemploymentlength (çalışma süresi) sütunlarında bulunan hatalı karakterler giderilmiştir.
- Kategorik veriler, sayısal verilere dönüştürülerek modelin işleyebileceği hale getirilmiştir.
- Modelde kullanılacak, krediyi etkileyebilecek en önemli özellikler (örneğin, yaş, gelir, kredi geçmişi uzunluğu vb.) belirlenmiştir.
- Veriler normalize edilmiştir.

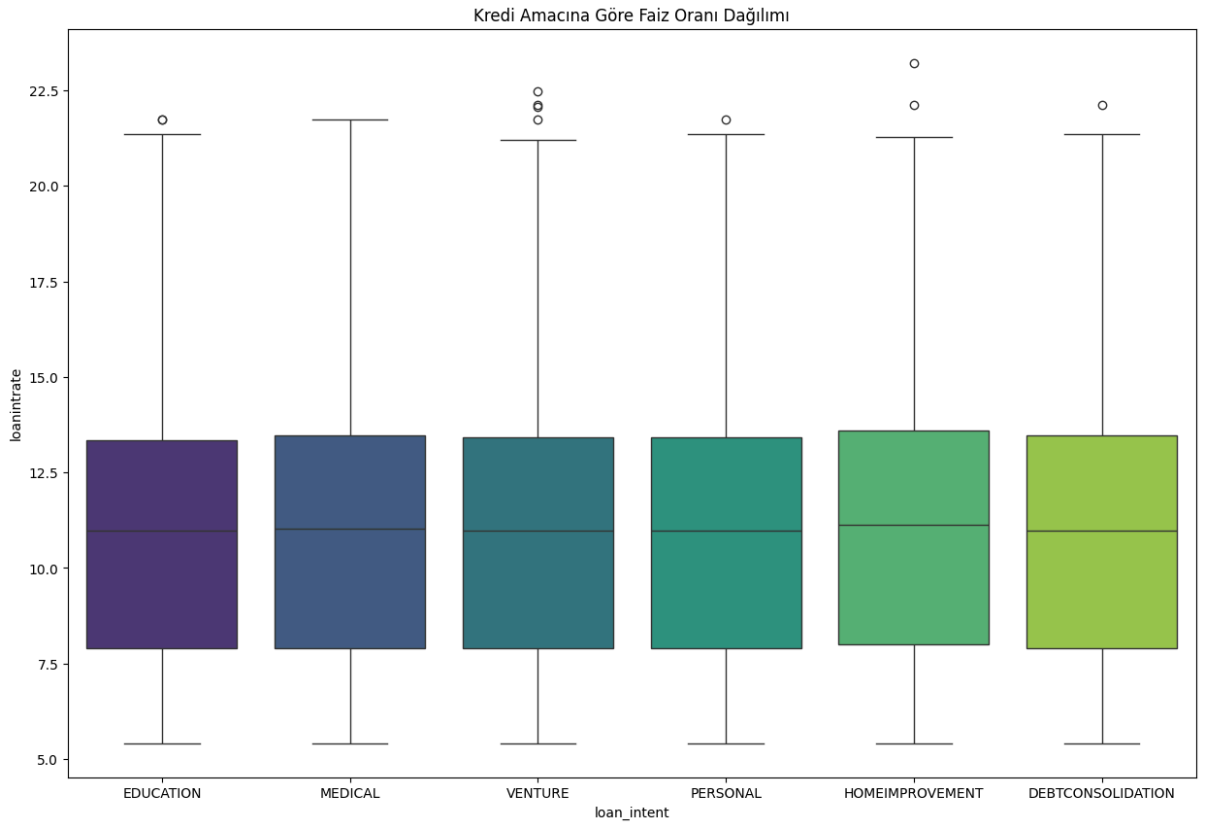
Keşifsel Veri Analizi

Aşağıdaki grafiklerde veri setimizin bağımlı ve bağımsız değişkenleriyle arasındaki ilişki görselleştirilmiştir. Bu grafiklere bakılarak veri setindeki attribute'lerin faiz oranı ve temerrüt olasılığı ile ilişkisi tespit edilmiştir.

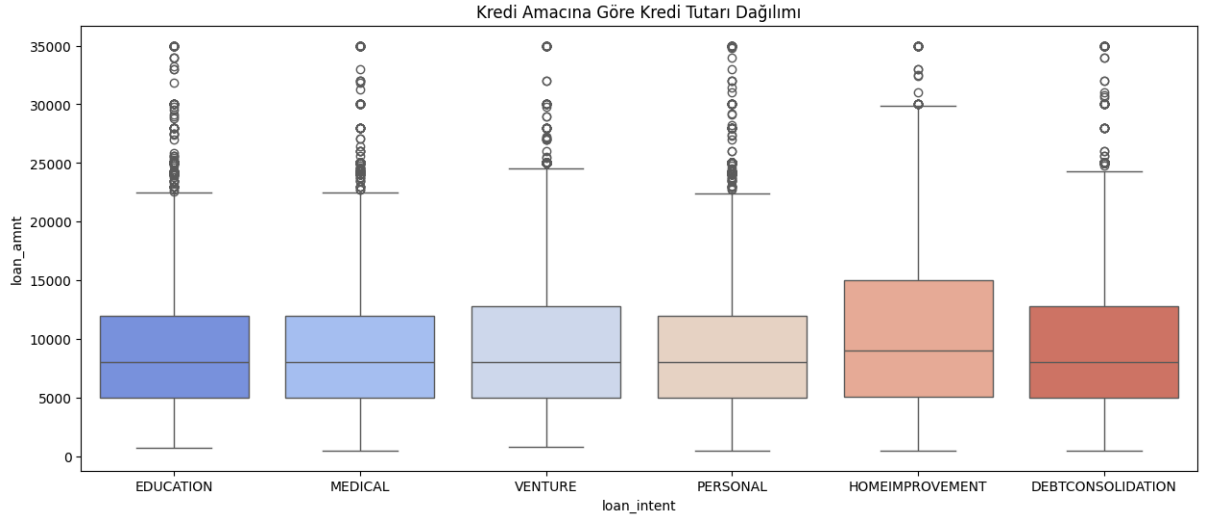




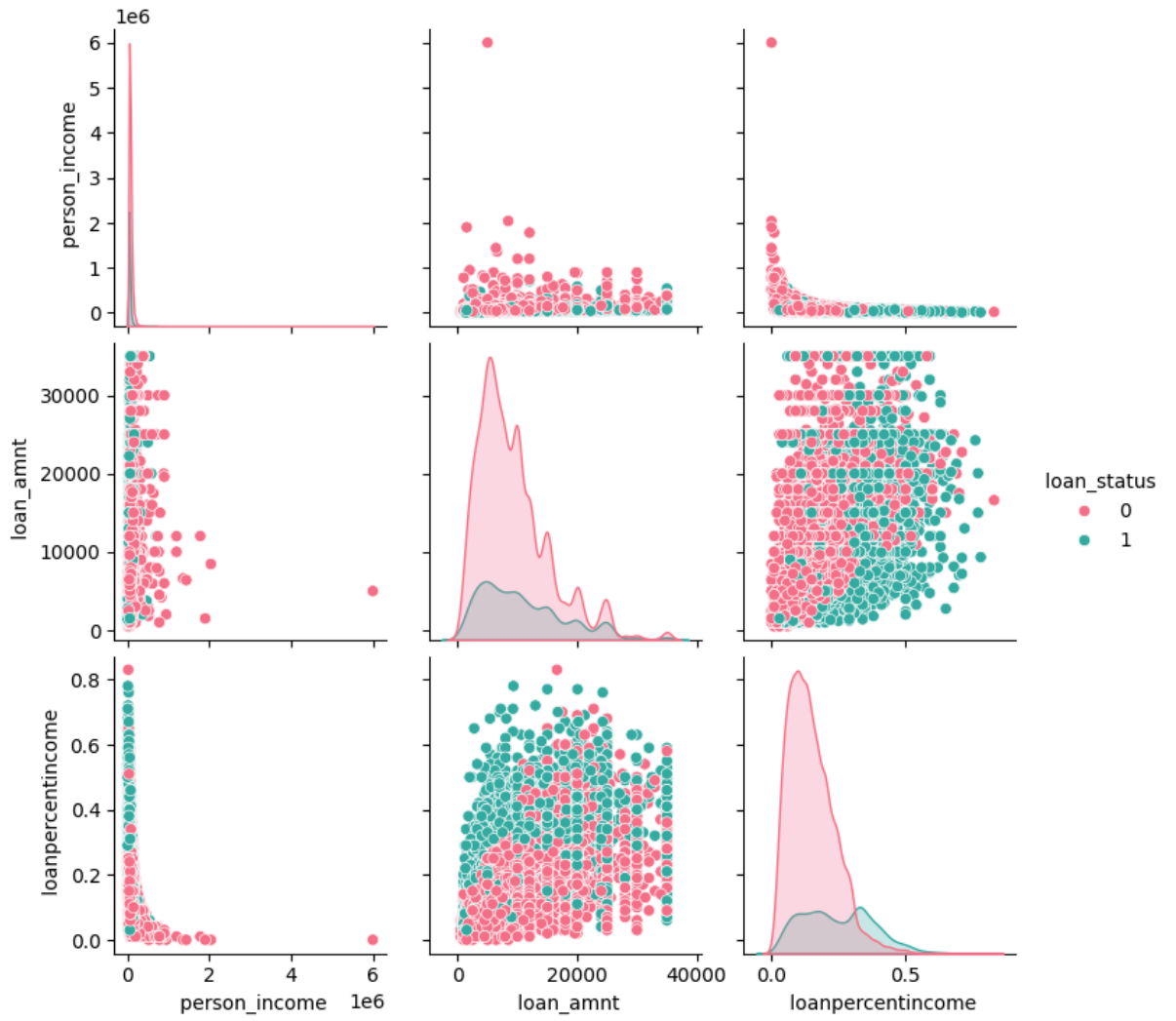
Grafik 3.



Grafik 4.



Grafik 5.



Grafik 6.

Model Oluřturma

Veri, eđitim (%80) ve test (%20) olarak b3l3nm3řt3ř. Modeli seęerken kredi temerr3t olasılıđını tahmini ięin 4 farklı sınıflandırma modeli kullanılmıřtır. Bu sınıflandırma modelleri, kredi temerr3t tahmini ięin uygulanmıřtır. Bu modeller arasında Logistic Regression, Decision Tree, Random Forest ve SVM yer almıřtır. Model deđerlendirmelerinde bařarı oranı (accuracy) en y3ksek olan model Random Forest olmuřtur. Bu model, kredi temerr3t riskini tahmin etme g3revinde daha dođru sonuęlar 3reterek, diđer modellere kıyasla 3řt3řn bir performans sergilemiřtir. Literat3ř taramasında da, Random Forest modelinin, finansal risk tahmini ve kredi skorlama gibi benzer alanlarda sıklıkla tercih edilen ve y3ksek dođruluk oranları elde eden bir model olduđu g3zlemlenmiřtir [3][4]. Bu nedenle, kredi temerr3t tahmininin g3venilir bir řekilde yapılabilmesi ięin Random Forest modelinin en uygun 33z3m olduđu sonucuna varılmıřtır.

Model Performansının Deđerlendirilmesi

Modelin bařarımı, dođruluk oranı (**accuracy**) ve F1 skoru, gibi metriklerle deđerlendirilmiřtir. Bu metrikler ařađıdaki g3rselde g3r3lmektedir.

G3rsel 1.

Model Performans Karřılařtırması:		
	Accuracy	F1 Score
Logistic Regression	0.874073	0.648768
Decision Tree	0.895692	0.757431
Random Forest	0.935774	0.826587
SVM	0.873599	0.648838

G3rsel 2.

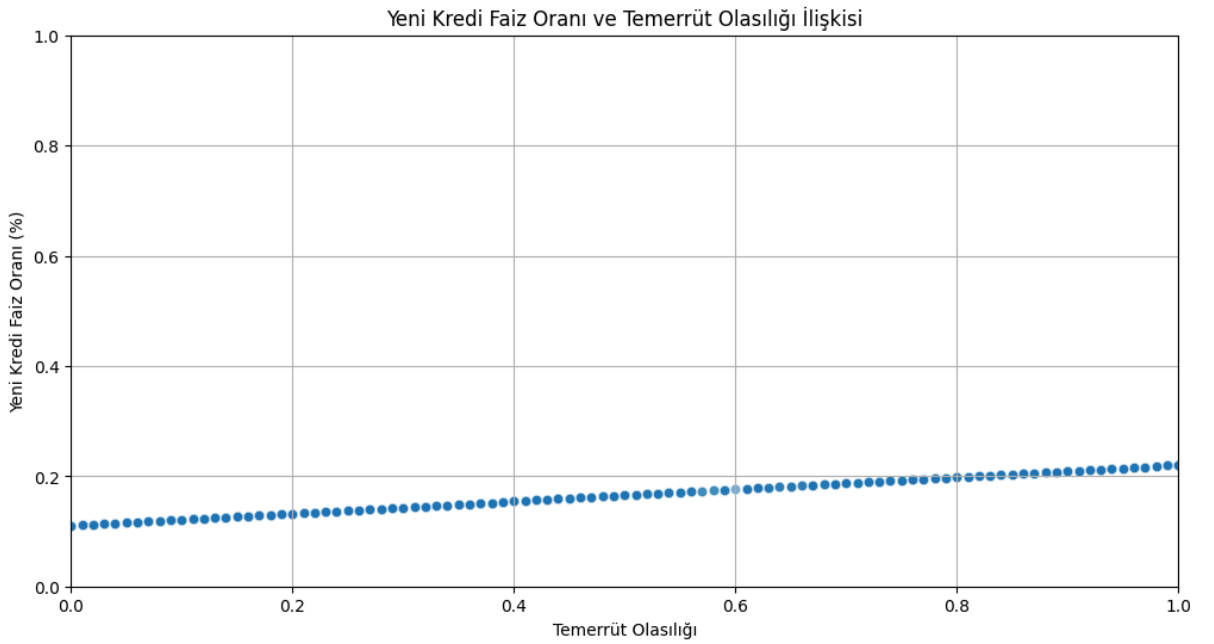
Accuracy: 0.94				
F1 Score: 0.83				
Classification Report:				
	precision	recall	f1-score	support
0	0.93	0.99	0.96	4991
1	0.97	0.72	0.83	1346
accuracy			0.94	6337
macro avg	0.95	0.86	0.89	6337
weighted avg	0.94	0.94	0.93	6337

Yukarıdaki görsel 1’de modelimizii derlerken denediğimiz 4 farklı algoritmanın değerleri görülmektedir. Bu değerlendirmeler göz önüne alınarak Random Forest kullanılmıştır. Görsel 2’de Kullanılan Random Forest algoritmasının detaylı değerleri görülmektedir.

Faiz Oranı Hesaplama:

Temel faiz oranı ve risk faktörü kullanılarak risk uyumlu faiz oranları hesaplanmıştır. Müşterinin temerrüt olasılığı ile doğru orantılı faiz oranları görselleştirilmiştir.

$$suggested_interest_rate = base_interest_rate + (risk_adjustment_factor \times predicted_default_probability)$$



Grafik 7.

Yukarıdaki grafikte ise train setinde kriterize edilmiş attribute’lerin kredi alabilme olasılığı değerlerinin veri setine göre modellenmiş halinin sonucu gösterilmiştir.

Bu sonuca göre yeni kredi faiz oranlarının, temerrüt olasılığına göre 10 ile 20 arası değiştiği grafikte görülmektedir. Temerrüt ve faiz oranının doğru orantılı olduğu ve modelin düzgün çalıştığını gözlemlenmektedir.

2. Web Uygulaması Geliştirme

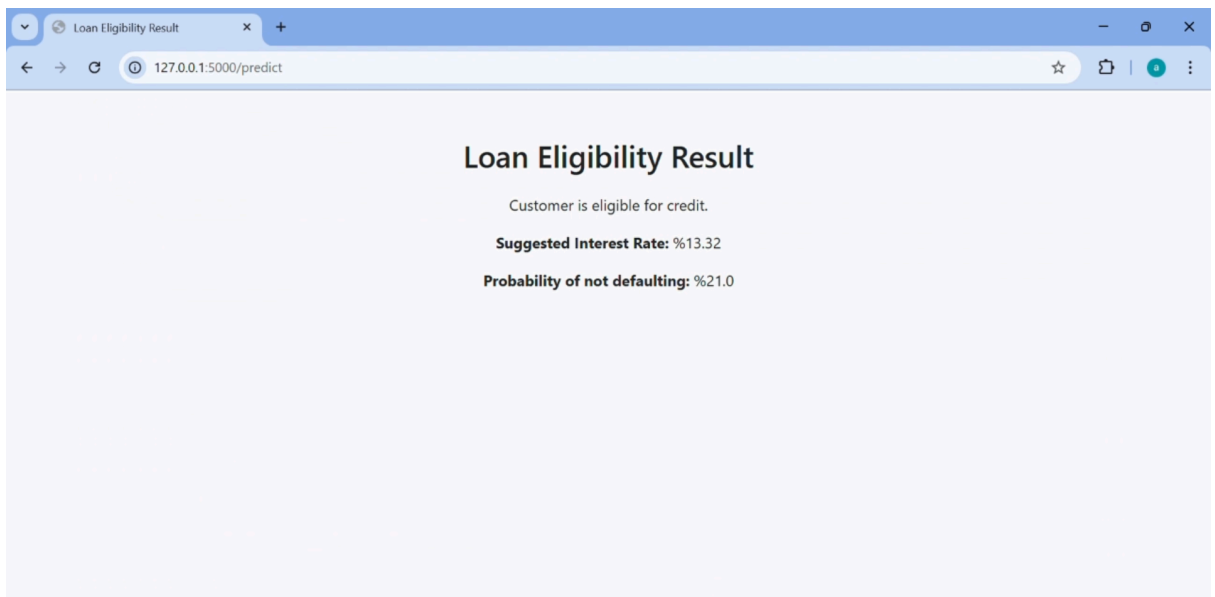
Kullanıcı girdilerine göre kredi onay-faiz oranlarını göstermek amacıyla bir web arayüzü tasarlanmıştır. Flask kullanılarak geliştirilen uygulama, kullanıcının sağladığı verilere göre kredi uygunluğunu değerlendirir ve buna göre önerilen faiz oranını sunmaktadır. Model, joblib ile eğitilip .pkl formatında kaydedilmiş olup, web uygulamasına entegre edilmiştir. Aşağıdaki görsellerde web sitemizin arayüzü gösterilmiştir.

Görsel 3.



The screenshot shows a web browser window with a single tab titled "Loan Eligibility Prediction". The address bar displays "127.0.0.1:5000". The main content area has a light purple background and a large heading "Check Your Loan Eligibility". Below the heading, there are six input fields arranged vertically, each with a label to its left: "Age:", "Income:", "Home Ownership:", "Employment Length (years):", and "Loan Intent:". The "Home Ownership:" field is a dropdown menu currently showing "Rent". The "Loan Intent:" field is also a dropdown menu currently showing "Education".

Görsel 4.



The screenshot shows a web browser window with a single tab titled "Loan Eligibility Result". The address bar displays "127.0.0.1:5000/predict". The main content area has a light purple background and a large heading "Loan Eligibility Result". Below the heading, there is a message "Customer is eligible for credit." followed by two lines of text: "Suggested Interest Rate: %13.32" and "Probability of not defaulting: %21.0".

Kaynakça

- [1] [Y. Wang, Y. Zhang, Y. Lu, and X. Yu, “A Comparative Assessment of Credit Risk Model Based on Machine Learning ——a case study of bank loan data,” *Procedia Computer Science*, vol. 174, pp. 141–149, Jan. 2020, doi: 10.1016/j.procs.2020.06.069.](#)
- [2] [https://www.openml.org/search?type=data&id=43454&sort=runs&status=active](#)
- [3] [R. Kurniawan, “Application of random Forest algorithm on credit risk analysis,” *Procedia Computer Science*, vol. 245, pp. 740–749, Jan. 2024, doi: 10.1016/j.procs.2024.10.300.](#)
- [4] [M. Madaan, A. Kumar, C. Keshri, R. Jain, and P. Nagrath, “Loan default prediction using decision trees and random forest: A comparative study,” *IOP Conference Series Materials Science and Engineering*, vol. 1022, no. 1, p. 012042, Jan. 2021, doi: 10.1088/1757-899x/1022/1/012042.](#)