**Data Dictionary**

The dataset consists of the following fields:

• Loan ID: A unique Identifier for the loan information.

• Customer ID: A unique identifier for the customer. Customers may have more than one loan.

• Loan Status: A categorical variable indicating if the loan was paid back or defaulted.

• Current Loan Amount: This is the loan amount that was either completely paid off, or the amount that was defaulted.

• Term: A categorical variable indicating if it is a short term or long term loan.

• Credit Score: A value between 0 and 800 indicating the riskiness of the borrowers credit history.

• Years in current job: A categorical variable indicating how many years the customer has been in their current job.

• Home Ownership: Categorical variable indicating home ownership. Values are "Rent", "Home Mortgage", and "Own". If the value is OWN, then the customer is a home owner with no mortgage

• Annual Income: The customer's annual income

• Purpose: A description of the purpose of the loan.

• Monthly Debt: The customer's monthly payment for their existing loans

• Years of Credit History: The years since the first entry in the customer’s credit history • Months since last delinquent: Months since the last loan delinquent payment

• Number of Open Accounts: The total number of open credit cards

• Number of Credit Problems: The number of credit problems in the customer records.

• Current Credit Balance: The current total debt for the customer

• Maximum Open Credit: The maximum credit limit for all credit sources.

• Bankruptcies: The number of bankruptcies

• Tax Liens: The number of tax liens.

**Business Objective**  
  
A financial institution wants us to help them identify customers who have a lesser chance of defaulting on the loan.  
  
The company management has asked the data science team to build a predictive model to identify who would be a good customer. Furthermore, they want the team to come up with questions to ask the client, based on the model, when they are applying for loan.  
  
**Data Understanding**  
The dataset resembles a real-world dataset and has many of the same challenges. It has:  
Missing values  
Spelling differences  
Punctuation format  
Duplicates rows  
  
**Data Preparation**  
Split your data into training and testing  
Start with Exploratory data analysis  
Data cleaning  
Handling the missing values  
Transform categorical data into numeric  
Feature Engineering (such as credit utilization)  
The goal is to clean the dataset and get it ready for the Algorithms  
Modeling  
Algorithm Selection  
Depending on the question at hand you can decide which algorithm  
to choose  
  
**Classification Question**  
Pick a classification algorithm  
Regression based  
Tree based  
Distance based  
Probability based  
Model Evaluation  
Evaluation criteria  
  
**Modeling:**  
Pick an algorithm  
Train the algorithm using training data  
Evaluate the trained model  
Use the trained model to predict who is a good customer  
on test data  
Come up with questions to ask the customer when they  
apply for a loan

**Veri Sözlüğü**  
  
Veri kümesi aşağıdaki alanlardan oluşur:

• Kredi Kimliği: Kredi bilgileri için benzersiz bir tanımlayıcı.

• Müşteri Kimliği: Müşteri için benzersiz bir tanımlayıcı. Müşteriler birden fazla krediye sahip olabilir.

• Kredi Durumu: Kredinin geri ödenip ödenmediğini veya temerrüde düşüp düşmediğini gösteren kategorik bir değişken.

• Mevcut Kredi Tutarı: Tamamen geri ödenen kredi tutarı veya temerrüde düşen tutardır.

• Vade: Kısa vadeli mi yoksa uzun vadeli mi olduğunu gösteren kategorik bir değişkendir.

• Kredi Puanı: Borçlunun kredi geçmişinin riskini gösteren 0 ile 800 arasında bir değerdir.

• Mevcut işteki çalışma süresi: Müşterinin mevcut işinde kaç yıldır çalıştığını gösteren kategorik bir değişkendir.

• Ev Sahipliği: Ev sahipliğini gösteren kategorik değişken. Değerler “Kira”, “Ev İpoteği” ve “Kendi” şeklindedir. Değer KENDİ ise, müşteri ipoteği olmayan bir ev sahibidir.

• Yıllık Gelir: Müşterinin yıllık geliri

• Amaç: Kredinin amacının açıklaması.

• Aylık Borç: Müşterinin mevcut kredileri için aylık ödemesi.

• Kredi Geçmişi Yılları: Müşterinin kredi geçmişindeki ilk girişten bu yana geçen yıllar. • Son gecikmeden bu yana geçen aylar: Son kredi gecikme ödemesinden bu yana geçen aylar.

• Açık Hesap Sayısı: Açık kredi kartlarının toplam sayısı

• Kredi Sorunları Sayısı: Müşteri kayıtlarındaki kredi sorunlarının sayısı.

• Mevcut Kredi Bakiyesi: Müşterinin mevcut toplam borcu

• Maksimum Açık Kredi: Tüm kredi kaynakları için maksimum kredi limiti.

• İflaslar: İflasların sayısı

• Vergi Hacizleri: Vergi hacizlerinin sayısı.

**İş Hedefi**

Bir finans kurumu, kredi ödemelerini geciktirme olasılığı daha düşük olan müşterileri belirlememize yardımcı olmamızı istiyor.

Şirket yönetimi, veri bilimi ekibinden iyi bir müşteri olabilecek kişileri belirlemek için bir tahmin modeli oluşturmasını istedi. Ayrıca, ekipten, müşterilerin kredi başvurusu yaparken modele dayalı olarak sorulacak sorular hazırlamasını istiyor.

**Verileri Anlama**

Veri kümesi, gerçek dünyadaki veri kümelerine benziyor ve birçok benzer zorluk içeriyor. Şu özelliklere sahip:

Eksik değerler

Yazım farklılıkları

Noktalama biçimi

Yinelenen satırlar

**Veri Hazırlama**

Verilerinizi eğitim ve test olarak ayırın.

Keşifsel veri analizi ile başlayın.

Verileri temizleyin.

Eksik değerleri işleyin.

Kategorik verileri sayısal verilere dönüştürün.

Özellik Mühendisliği (kredi kullanımı gibi)

Amaç, veri kümesini temizlemek ve algoritmalar için hazır hale getirmektir.

Modelleme

Algoritma Seçimi

Ele alınan soruya bağlı olarak, hangi algoritmayı seçeceğinize karar verebilirsiniz.

**Sınıflandırma Sorusu**

Bir sınıflandırma algoritması seçin.

Regresyon tabanlı

Ağaç tabanlı

Mesafe tabanlı

Olasılık tabanlı

Model Değerlendirme

Değerlendirme kriterleri

**Modelleme**

Bir algoritma seçin.

Eğitim verilerini kullanarak algoritmayı eğitin.

Eğitimli modeli değerlendirin.

Eğitimli modeli kullanarak test verilerinde

iyi bir müşteri olan kişileri tahmin edin.

Müşteriler kredi başvurusunda bulunduklarında

sorulacak soruları hazırlayın.