Project Summary

Problem Statement

The loan companies have various customer, some of them is paid off, or defaulted. Based on different data and information about the customers, can we tell which customer will be paid off or go default? Which machine learning algorithms will do the best job to describe the dataset?

Solution

This is a binary classification problem, whether the customer is either paid off or defaulted. We want to find a model which can fit the data with the least amount of errors. The model that going to be tested in this project are KNN, Decision Tree, SVM, and Logistic Regression. At the end, it will evaluate each model and put themes side by side. The best one to compare them is to use a confusion matrix, look at the true positive, false negative, etc. F1 score is a single value that can summarize the confusion matrix.

Data Source

The data set can describe in the following table:

Field	Description
Loan_status	Whether a loan is paid off on in collection
Principal	Basic principal loan amount at the
Terms	Origination terms which can be weekly (7 days), biweekly, and monthly payoff schedule
Effective_date	When the loan got originated and took effects
Due_date	Since it's one-time payoff schedule, each loan has one single due date
Age	Age of applicant
Education	Education of applicant
Gender	The gender of applicant

There are 400 rows of data in this data set.

Programing Environment

Window 10/11, Python 3.9, Jupyter Lab, Google Chrome

Project Summary

Libraries Used

Pandas, Sklearn, Matplotlib, Numpy, Itertools

Result

The most important finding in this project, can be summarized in this chart.

Algorithm	Rank	Jaccard	F1 Score	Log Loss
KNN	#1	0.51	0.63	11.51
Decision Tree	#2	0.56	0.70	10.23
SVM	#3	0.64	0.75	7.03
Logistic Regression	#4	0.56	0.66	8.95

There isn't a thing called a perfect model, but this data set, K-Nearest Neighbor seems to be the best solution, if we compare base F1 Score value.

Project Link and Source Code

Project in IPYNB:

https://github.com/GiggleSamurai/Feature-Engineering-and-Classification-ML-on-Customer-Loan/blob/935a5fee9c1ac6dcdd0f005314b3c1cb8d599d1b/Class%20Loan%20ML%20Project.ipynb