

ABSTRACT

This article provides a brief overview of the integration of machine learning into credit card approval. Modern applications face issues such as scalability and subjective judgment that machine learning aims to solve. We discuss commonly used algorithms, case studies, and ethical considerations. Future directions include advanced machine learning techniques and advanced model interpretation to improve decision-making capabilities. This article takes an in-depth look at how machine learning can revolutionize the credit card approval process, reduce financial risk, and improve customer satisfaction.

INTRODUCTION

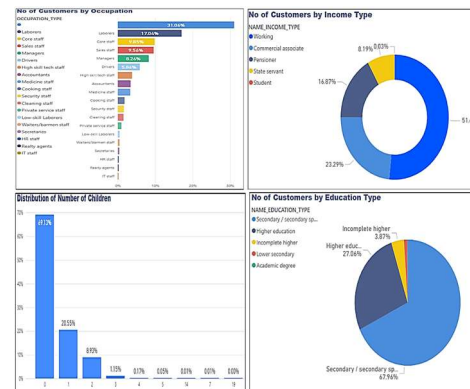
Credit card approval is an important part of finance today and facilitates access to credit while managing risk. Traditional methods are often unreliable and error-prone. Blended machine learning (ML) offers the promise of decision-making and analyzing big data to accurately measure creditworthiness. This introduction provides an overview of the evolution of credit card authorization, highlighting the benefits machine learning brings in terms of efficiency, accuracy, and integrity.

Machine learning algorithms can analyze large amounts of data to identify patterns in the credit card approval process and make predictions for decision-making. Machine learning models can create greater credit and personality assessments of applicants using different types of data, such as financial history and behavior.



Dataset & Methodology

As shown in 1 graph Distribution of Numbers of Children, 69 % are not have a child, 20% have one child, 9% have two children and 1% have 3 children. According to the 2 graph No of Customers by family status 69% are married, 13% are single, 5% are separated and 4% are widows. As shown in figure 3 No of Customers by Education Type figure 68% have secondary education, 27% are have higher education and 4% are have incomplete higher education. According to the graph 4.4 No of Customers by Income Type, 62% are working, 23% are commercial associates, 16% are pensioner, 8% State servant.



Graphical and numerical representation of data provide better insight about particular data set. Graphical representation of our data set is described below.

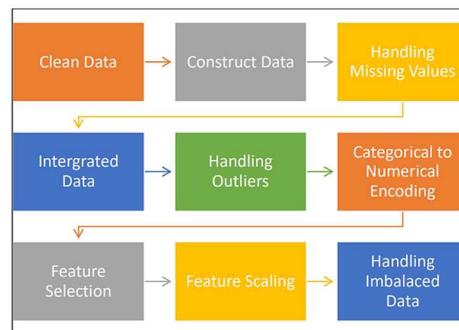


Fig : Proposed Methodology

Data preparation phase, which is often referred to as "data munging" or "Data Preprocessing" prepares the final data set(s) for modeling. Python programming with libraries /packages use to prepare the data set. Key main areas related to data preparation phase considered in the project as follows. Data Preparation with Explanatory Data Analysis (EDA) under each preparation activity .Feature Selection from finally prepared data set.

Methodology Cont..

Neural Network (NN) is evolved from biological neural network of human brain. It is deep learning algorithm and use as information processing technique. We can use NN not only for a classification problem but also regression. Neural network may contain 3 layers as follows: Model Evaluation Data Set Testing Data Set Testing Model Learning Algorithm - Model (NN) Data Preparation Feature Selection Feature Scaling Training Data Set Training Model • Input Layers - Raw information feed as input to the network • Hidden Layer - Input unit and weight. There can be many hidden layers. • Output Layer - This layer depends on hidden layer and weights or input layer. Prediction related to response variable return in output layer. Two layers using ReLU and one more layer using Sigmoid.

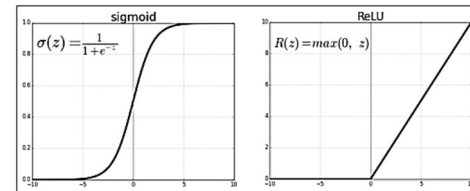
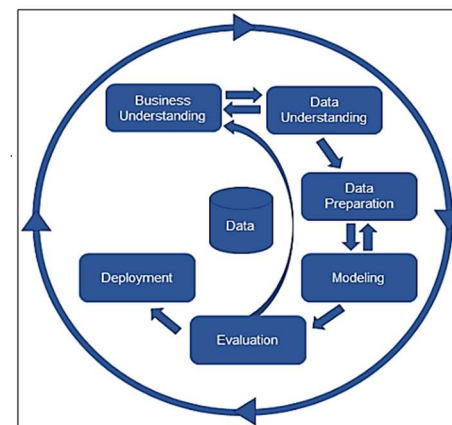


Figure :Sigmoid and ReLU activate Functions

Business Understanding - Understanding of objectives and requirements and produce of detail plan for project focus in here. Data Understanding - Focusing on identify, collect and analyze data. Data format/fields identification, identify relationships by visualization; verify data quality (clean/dirty) are the main activities carried during this phase. Modeling - Determine selection of algorithms, generate test design, build model and asses model are main activities carried out in this phase. Evaluation - Focusing on identification of which model best fit the business requirement. By determining whether to proceed to deployment or iterate further will be judge in here. Deployment - Focusing on accessible methods for developed model output/results. Deployment plan, monitoring and maintenance.

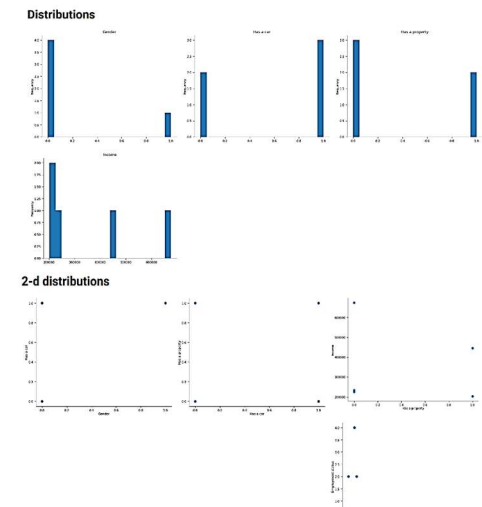


Tools and Technology

We have implement this project by using Microsoft Power BI for Data Visualization and Python with Spider (as Scientific Python Development Environment) for data programming. There were several python libraries being used for analysing the data, model building and model validation. They are Pandas, NumPy, Matplotlib, Scikit-Learn, keras .

- Pandas :Mainly used for data analysis, import data from csv files, data manipulation such as selecting, merging, reshaping and for data cleaning process.
- NumPy: Used for working with n-dimensional arrays and linear algebra.
- Matplotlib :Used for creating static and interactive visualization in data.
- Keras Used for implementing NN machine learning algorithm (from keras.models import Sequential, from keras.layers import Dense).

Result & Conclusion



We have obtained the publically available data set and explanatory analysis was carried out to understand the data set. Then conducted several activities related to data preparations such as data preprocessing, feature selections and feature scaling. To achieve a desired outcome, it is very important to carry out these activities accurately. We have divided the data set into two parts as a training and test data set and the intended purpose is to validate the accuracy of the model. NN model performances tested using low and high learning rates. Accuracy is 0.78, Precision is 0.81, Recall is 0.73 and AUC is 0.79 with a higher learning rate of 0.01. Accuracy is 0.76, Precision is 0.76, Recall is 0.74 and AUC is 0.85 with a lower learning rate of 0.001. Precision and recall values are high in higher learning rate. Smaller batch size provides a slower learning process. However, a small learning rate gave better AUC at 0.85 for NN compared to high learning rate.