PROJECT 1 REPORT

BY KATLEGO LETSIDIDI

Alpha dreamers Banking consortium is a financial institution that provides business loans, savings accounts, and checking accounts for individuals and businesses. In their annual review, the board of directors for the institution saw it fit to enhance its internal operations to stay ahead of the competition. A new wing of business intelligence has been initiated to assist the loans department to automate, predict and formalize the business processes. The loans department has realized that there is a high defaulter rate for personal loans and measures must be put in place to minimize the defaulter rates.

As a technician in the BI Department, we have been tasked to develop a machine learning algorithm that will be able to predict the credibility or risk factor of any future loan applicants using data previously collected from past applicants. The original data includes several of the applicant’s details that might share a correlation with an applicant’s risk factor. These are; Income, Age, House ownership status, Car ownership status, Profession, State in which the applicant lives, City in which the applicant lives and whether they have defaulted on loans.

Part of the task was to determine which, if any, of these factors have an influence on whether an applicant is a risk or not. We used different charts to properly analyse the data. The insights we were able to gleam from the data are as follows:

1. The relationship between marital status and risk of defaulting

From the data we were able to realise that single people are more likely to default from loan payment then married couples

1. The relationship between car ownership and risk

From the data we were able to surmise that applicants that own cars are less likely to default from loan payments than applicants who do not.

1. The relationship between house ownership and risk of defaulting

It is shown that people who live in rented houses are most likely to default from loan payments and people who neither own houses nor rent them comes right after. People who own the houses they’ve settled in are the least likely to default from paying of their loans

1. The relationship between age and risk of defaulting

We used a line graph to visualize the relationship between age and the risk of defaulting. From the data it is shown that younger people are most likely to default from their loan payments, the highest risk age group being that of 20 – 25.

1. The correlation between income of an applicant and their risk of defaulting

A line graph was used to represent the relationship between an applicant’s income and the average default rate. It was found that the risk of defaulting reduces with an increase in income. Applicants who earn more are less likely to default from loan payments than those of lower incomes

1. The relationship between profession and risk of defaulting

A bar chart was used to represent this relationship because of the categorical nature of the data. It was found that applicants in certain professions are more likely to default from loan payments such as police officers, whilst applicants in other professions were less likely to default from their loan payments such as petroleum engineers. We were also able to deduce a relation between these jobs and the income they produce thus reinforcing the previous point on the nature of the relationship between income and risk.

1. The relationship between years of working experience and risk of defaulting

A scatter graph was used to find the correlation between the amount of working experience an applicant has and their risk of defaulting from loan payments. It was found that applicants with more work experience are less likely to default from loan payments than those who have less.

1. The relationship between the amount of time a person has been working at their current job and risk of defaulting

A scatter graph was also used to find the correlation between the amount of time an applicant has been working at their current job and their rate of defaulting from loan payments. On average people who have less time at their current job run a higher risk of defaulting, but this is only a weak correlation, and its actual influence is negligible.

1. The relationship between the amount of time an individual has been occupying their current house (whether owned or rented) and risk of defaulting

A line graph was used to determine the nature of the relationship between risk of defaulting from loan payments and the amount of time an applicant has been occupying their current house. It was found that the longer an applicant has been occupying their current house the less of a risk the pose. There is a significant drop in risk after an applicant has been living in the same house for more than eleven years

1. The relationship between the state an applicant lives in and their risk of defaulting

A bar chart was used to determine the relationship between the state in which an applicant lives and their likelihood of defaulting from their loan payments. It was found that people living in certain states, such as Manipur, have a higher risk of defaulting from their loan payments than other states.