CREDIT CARD APPROVAL MODEL

By: Alec Hing

Outline

- Overview & Business Problem
- The Data Involved
- The Approach and Methods
- Results
- Improvements
- Conclusion

Overview & Business Problem

- The objective of this project is to determine if a person applying for a credit card is a risky applicant or not
- It can be challenging for a person to analyze all the factors in another person’s life and decide if they are eligible for a credit card
- Everyone’s background is different so there is no one feature someone can look at and make this decision

The Data Involved

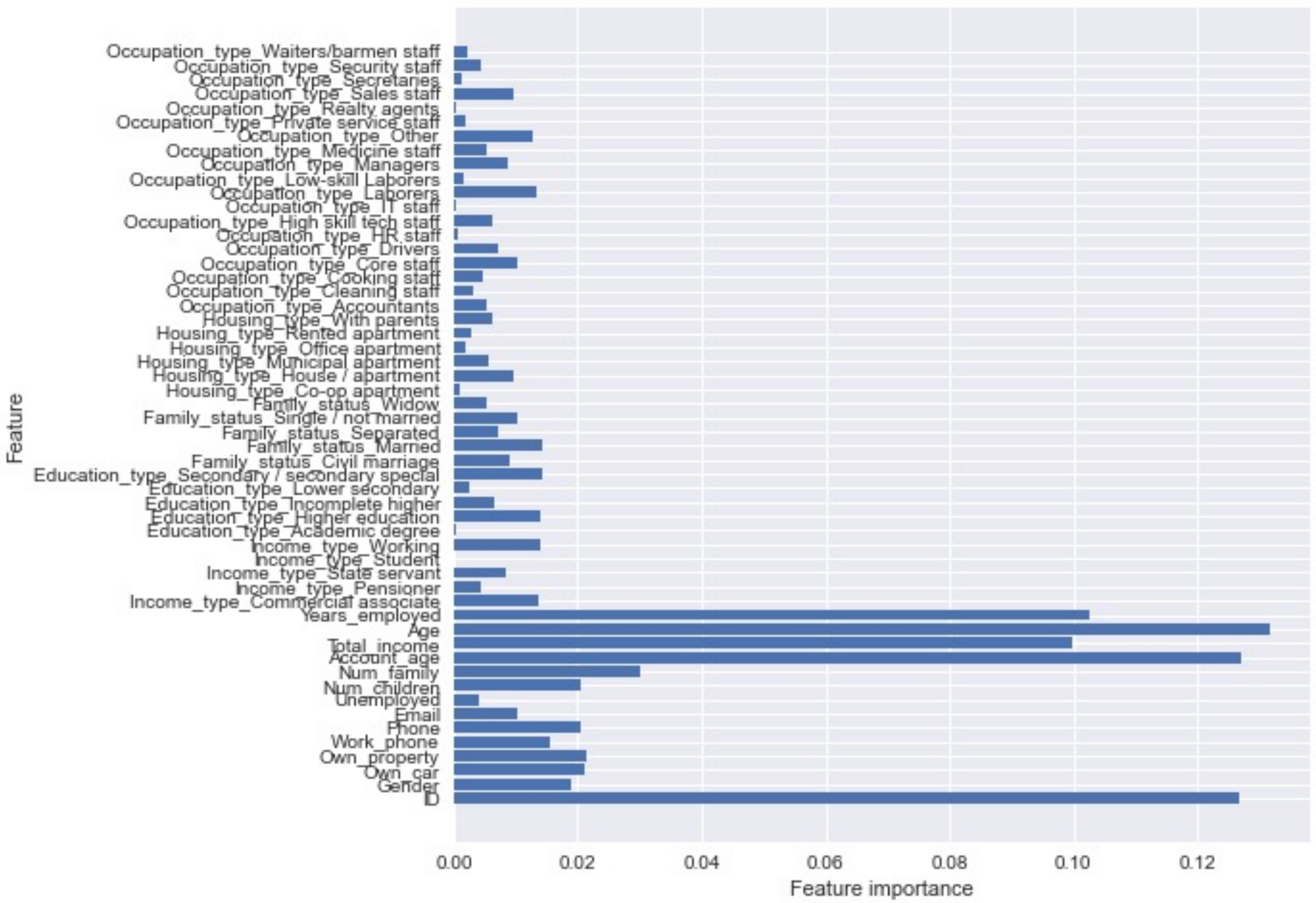
- Data analyzed in this analysis is from Kaggle
- The main features considered were Age, Total Income, Years Employed, Account age, and Income Type
- An at-risk applicant was decided to be someone who has not paid their bills in more than 30 days
- Features that did not play an impactful role in the results were occupation type, whether they owned a car or property, family status, education level, or gender

Data (Continued)

- One model created in this analysis was able to have an accuracy of 74%, meaning it was able to predict 74% correctly out all observations
- This model was not chosen because a majority of the 74% were those who were safe applicants
- Another model was able to predict better than any other model those individuals that were a risky applicant at 54%
- For reference, all other models besides this one was under 20%

Approach & Methods

- An analysis of the most important features to the chosen was made (See next slide)
- Many models were created to determine the best result using numerous machine learning techniques (XG Boost, Grid Search CV, and Random Forests) to name a few
- A final model was chosen by prioritizing the model that predicted the most at risk applicants for a credit card



Occupation_type_Waiters/barmen staff
Occupation_type_Security staff
Occupation_type_Secretaries
Occupation_type_Sales staff
Occupation_type_Realty agents
Occupation_type_Private service staff
Occupation_type_Other
Occupation_type_Medicine staff
Occupation_type_Managers
Occupation_type_Low-skill Laborers
Occupation_type_Laborers
Occupation_type_IT staff
Occupation_type_High skill tech staff
Occupation_type_HR staff
Occupation_type_Drivers
Occupation_type_Core staff
Occupation_type_Cooking staff
Occupation_type_Cleaning staff
Occupation_type_Accountants
Housing_type_With parents
Housing_type_Rented apartment
Housing_type_Office apartment
Housing_type_Municipal apartment
Housing_type_House / apartment
Housing_type_Co-op apartment
Family_status_Widow
Family_status_Single / not married
Family_status_Separated
Family_status_Married
Family_status_Civil marriage
Education_type_Secondary / secondary special
Education_type_Lower secondary
Education_type_Incomplete higher
Education_type_Higher education
Education_type_Academic degree
Income_type_Working
Income_type_Student
Income_type_State servant
Income_type_Pensioner
Income_type_Commercial associate
Years_employed
Age
Total_income
Account_age
Num_family
Num_children
Unemployed
Email
Phone
Work_phone
Own_property
Own_car
Gender
ID
Feature
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Feature importance

Results

- 7 different models were created to determine the best model
- 6/7 models created had subpar (under 20%) results when it came to predicting actual risky applicants of credit cards
- Only one of them predicted 54% correctly of risky applicants
- The features that have the greatest impact to the results of the model chosen are age, income, account age, and employment length

Improvements

- One improvement that can be done is to try improving the over all ability of the other 6 models to be able to predict more risky applicants
- A possible solution to this is fine tuning of the parameters used
- Another improvement would be to remove features not needed in this analysis (whether they had an email, phone, or work phone) to name a few

Conclusion

- The recommendation would be to go with the model that predicts more risky applicants than those who are responsible spenders
- It would hurt a business more if the model said a person was a safe buyer when really, they were not
- Data to prioritize gathering from each applicant would be their age, account history, total income, and how long they’ve been employed for

Thank You

Email: ahing619@gmail.com

GitHub: @ahing

LinkedIn: https://www.linkedin.com/in/alec-hing