Instant Credit Card Approval

Shitij Seth YuanJea Hew Muhammad Salim Deryck Ma

Project Description

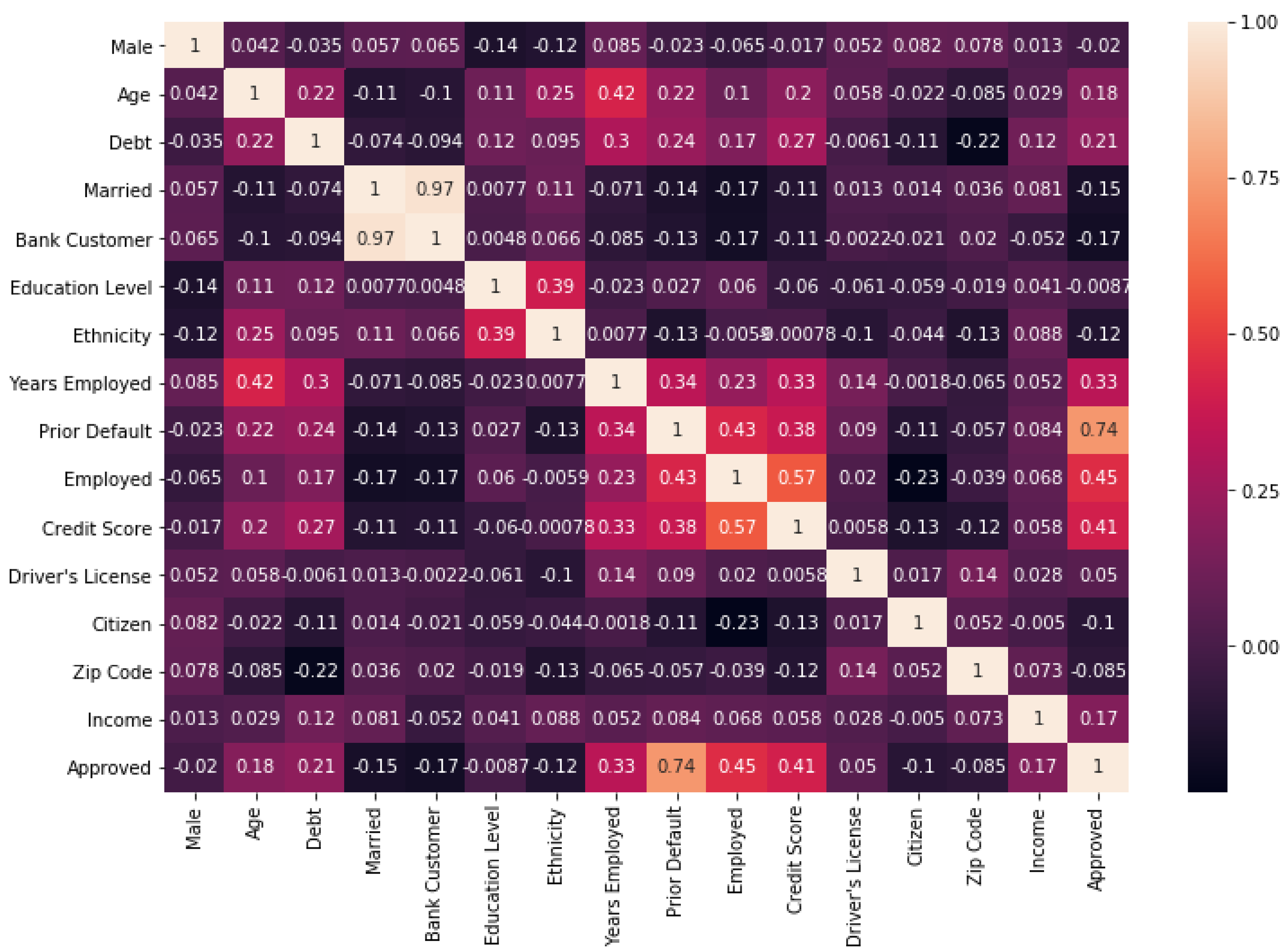
- As more and more people are moving towards digital cashless methods of payment, credit cards have been the top choice of consumers in the US which results in tons of applications for banks to make decisions on.
- The ultimate goal of the project was to build an algorithm which predicts a decision for an application instantly and one which can actually be used by a bank in the real world.
- The scope of this project is to:
  - (a) enhance results from the refence paper [1].
  - (b) compare the accuracy from different algorithms.
  - (c) determine the features that most affect card approval.

Challenges

- A number of supervised machine learning algorithms were analyzed in approaching the project goals. Each had different pros and cons.
- For instance, SVM provides high accuracy but takes a long time to run.
- Therefore, developing an algorithm which provides the **best combination of accuracy and run time** was the biggest challenge.

Potential Applications

- Banks can utilize machine learning algorithms to analyze customer data to decide credit card applications.
- These algorithms can be used to assess any type of credit applications such as loans.



Data Analysis

Feature : [Values] → **[Data Transformation]**

- Male: a,b → 0,1
- Married: u,y,l → 0,1,2
- Bank Customer: g,p → 0,1
- Education: c,d,i,j,k,m,r,q,w,x,e,aa,ff → 0,1,2,3,4,5,6,7,8,9,10,11,12
- Ethnicity: v,h,bb,j,n,z,dd,ff,o → 0,1,2,3,4,5,6,7,8
- Prior Default: t,f → 1,0
- Employed: t,f → 1,0
- Drivers License: t,f → 1,0
- Citizen: g,p,s → 0,1,2
- Approval: +,- → 1,0

	Age	Debt	Year Employed	Credit Score	Income
Mean	31.50	4.83	2.24	2.50	1013.76
Median	28.42	2.84	1.00	0.00	5.00
Max	76.75	28.0	28.50	67.00	2000.00
Min	13.75	0.00	0.00	0.00	0.00

Methodology

- **Artificial Neural Network (ANN)**
  Data points interconnected in layers forming a network similar to ones inside the human brain
- **Logistic Regression (LR)**
  Maximizes the likelihood to find best parameters and predicts using the logistic function
- **Support Vector Machine (SVM)**
  Fits multi-dimensional non-linear planes to create decision boundary
- **Decision Trees (DT)**
  Uses a bunch of yes-no question to classify
- **Random Forest (RF)**
  Uses the majority vote from a combination of several decision trees
- **K-Nearest Neighbors (KNN)**
  Selects the *k* number of nearest neighbors surrounding a query point and picks majority vote as its prediction
- **Ensembling**
  Uses the majority vote from ensembled algorithms
- **Grid Search**
  Exhaustive search of the parameter space to find optimum hyperparameter values was used for all algorithms using 85-15 as the train-test split

Glossary of Technical Terms

Logistic function: Common S-shaped curve used to bring non-linearity to predictions

Train-test split: Data is segmented into training and testing set based on this percentage divide

Results

- Laws prohibit credit discrimination on the basis of race, color, religion, national origin, sex, marital status and age [2].
- Only top 4 most important features used based on the correlation matrix (bottom left).
- **Individual algorithm comparison:**

Algorithm	Accuracy (%)	Run Time (s)
ANN	92.96	0.035865
LR	93	0.004508
SVM	88	0.010412
KNN	93	0.004109
DT	91	0.222498
RF	90.56	0.001875

- **Best model**: Ensemble of LR, RF and DT with 94% accuracy
- Our model achieved a higher accuracy (94%) than that of the model used in the reference paper (84%).
- **Comparison from baselines:**

	Accuracy (%)	Run Time (s)
Data Baseline	55	N/A
Default Parameter Baseline	86.52	0.0628
Best Model	94	0.0539

References

1. Kuhn. Analysis of Credit Approval Data. *Internet* http://rstudio-pubs-static.s3.amazonaws.com/73039_9946de135c0a49daa7a0a9eda4a67a72.html
2. "Your Equal Credit Opportunity Rights." Consumer Information, 13 Mar. 2018, www.consumer.ftc.gov/articles/0347-your-equal-credit-opportunity-rights.

Acknowledgments

This project was mentored by Dr. Christina Duron, whose help is acknowledged with great appreciation.
Support from our instructor, Dr. Calvin Zhang-Molina is also gratefully acknowledged.