Title: Credit Score Evaluation of Customer Using Machine Learning

SOMAIYA VIDYAVIHAR UNIVERSITY
K J Somaiya College of Engineering

Somaiya TRUST

Problem Definition

- Credit Scoring is the one of the most important and critical sources of analysis to the financial market of lending decisions and profitability.
- With the continuous development of the financial and banking institutions, credit products play an increasingly important role in the economy.
- These makes an Individual and any banking or Lending Institution to identify its Customer Credit Worthiness and Market Credibility.
- Statistical evaluation of Individual's Financial behavior have resulted in various fault and thus resulting to loss.
- There's a need to develop a strong System which can identify the Customer's financial behavior using Machine Learning Algorithms.
- We aim to develop a strong Credit Score calculation System using Machine Learning algorithm which will help user to identify its credit worthiness and market credibility and predict whether an individual is eligible for a particular Loan or not.

SRS(FUNCTIONAL & NON-FUNCTIONAL REQUIREMENTS)

Functional Requirements

- User Login and Registration
- Credit Score Calculation
- Investment Options
- Loan Approval & EMI Calculation

Non Functional Requirements

- Reliability
- Scalability
- Availability
- Security
- Portability
- Maintainability

Literature Review

Papers	Author	Methodology	Attributes	Accuracy
Credit Scoring Using Machine Learning Techniques -IJCA Volume 161-11,March 2017	Sunil Bhatia, Pratik Sharma, Rohit Burman	Customer Behavior as a Good(1) or bad(0) is classified using Logistic Discriminant Analysis, Artificial Neural Network, XGBoost, SVM	Gender, Married, Income, Education, Dependents, Previous Loan ,House(own/rented),credit card, Number of Previous loans, Jobtype	LDA-0.73 ANN-0. 82 SVM-0. 84
Credit Score using Machine Learning combining social network information . -December 2019	Beibeiniu, Jinzheng Ren and Xiaotao Li	Credit score of a customer is classified using its Social Networking information considering peer to peer lending, online microlending market etc. using Random Forest, Adaboost, Light GBM	Social Stability, Social Exposure, Social Quality, Gender, Age, Income, Education ,Job, Years at Residence, Credit History, City, Dependants , type of loan	Random Forest-0.74 Adaboost-0.78
Credit Scoring Model using ML – March 2018	Jasmina Nalic, Amar Svraka	Credit Score is calculated as Good or a bad Customer using Oracle Data Miner, ROC Curves and Graphical analysis	Office Details, other loans, Credit history , Nationality , Current address, income, dependents, Education ,Gender, Marriage, Self-Employed, DOB	LR-0.82 SVM-0.79 ANN-0.76

Literature Review

Papers	Author	Methodology	Attributes	Accuracy
Credit Score Model using data Mining	Yap Beewah, Irma Rohaiza Ibrahim	Customer financial Behavior as a Good or a bad using Machine Learning Algorithms like Neural Network, Logistic Regression ,Decision Tree, SVM	Age, Gender , Income, Dependents, Current Address ,Married , SelfEmployed , City, Previous Loan ,Loan Period, Job Title , House(own/Rented)	ANN-0.80 LR-0.82 Decision Tree-0.78 SVM-0.79
Credit Score Using Naïve Bayesian Approach	Olatunj J Okesola, kennedy O okokpuji	Credit score using naïve bayes with probabilistic value and predicting the customer as a Good or Bad	Age, Gender, Marriage, Income, Dependents, Jobtitle, SelfEmployed, Previous Loan ,House(own/Rented),City, Monthly Expense	Naïve Bayes-0.86
Credit Risk Prediction Using Data Mining Algorithm	Archana Gahlaut, Prince Kumar Singh	Customer behavior as a Good or a bad is calculated using machine learning tools like Neural Network, Decision tree, Adaptive boosting, SVM, Logistic Regression	Gender, Marriage, Income, Dependents, Property Details, House(own/rented), type of loan, City, Monthly expense ,previous loans, dependent income, savings, credit history	ANN-0.82 Decision tree-0.78 Adaptive boosting-0.76 SVM-0.84 LR-0.88

Technologies Used

- **Languages Used**

HTML,CSS,JavaScript,Bootstrap,Java,Python3,PHP

- **Tools Used**

Xampp Control Panel, Sublime Text , Eclipse, Jupyter Notebook, Microsoft Azure ,Microsoft Excel ,Star UML, cmd

- **Libraries Used**

Pandas, Numpy, sklearn,Flask

ImplementationDetails

Attributes	Values
Gender	Male/Female
Married	Yes/No
Dependents	0-5
Education	Graduate/Not Graduate
Self Employed	Yes/No
Applicant Income	Integer Value
CoApplicant Income	Integer Value
Previous Loan	Yes/No
Previous Loan Amount	Integer
Previous Loan Period	Number of Months
Property	Urban/Semi Urban/Rural
Monthly Debt Ratio (Monthly Living Cost/Monthly Gross Income)	Integer Value

CONDITION	CASES	POINTS	Percentage
PAYMENT HISTORY(30%)	Standard/No Overdue	300	100%
Total points=300	Overdue 30 days	225	75%
	Overdue 60 days	150	50%
	>90 days	0	
LOANS(15%)	1 loan	60	40%
Total points=150	2 loans	105	70%
	3 loans	120	80%
	4-5loans	127.5	85%
	>5loans	97.5	65%
LENGTH OF CREDIT(15%)	<1yr	45	30%
Total points=150	>1yr-2yr	60	40%
	>2yr-3yr	120	80%
	>3yr-5yr	135	90%
	>5yrs	150	100%
TYPE OF CREDIT(10%)	House Loan	70	70%
Total points=100	Personal Loan	80	80%
	Business Loan	90	90%
OTHERS(30%)	...		
Total points=300	Credit Card		
	Veto Power		
	Etc.		

MAX SCORE	1000
HIGH RISK	<400
GOOD CUSTOMER	>750

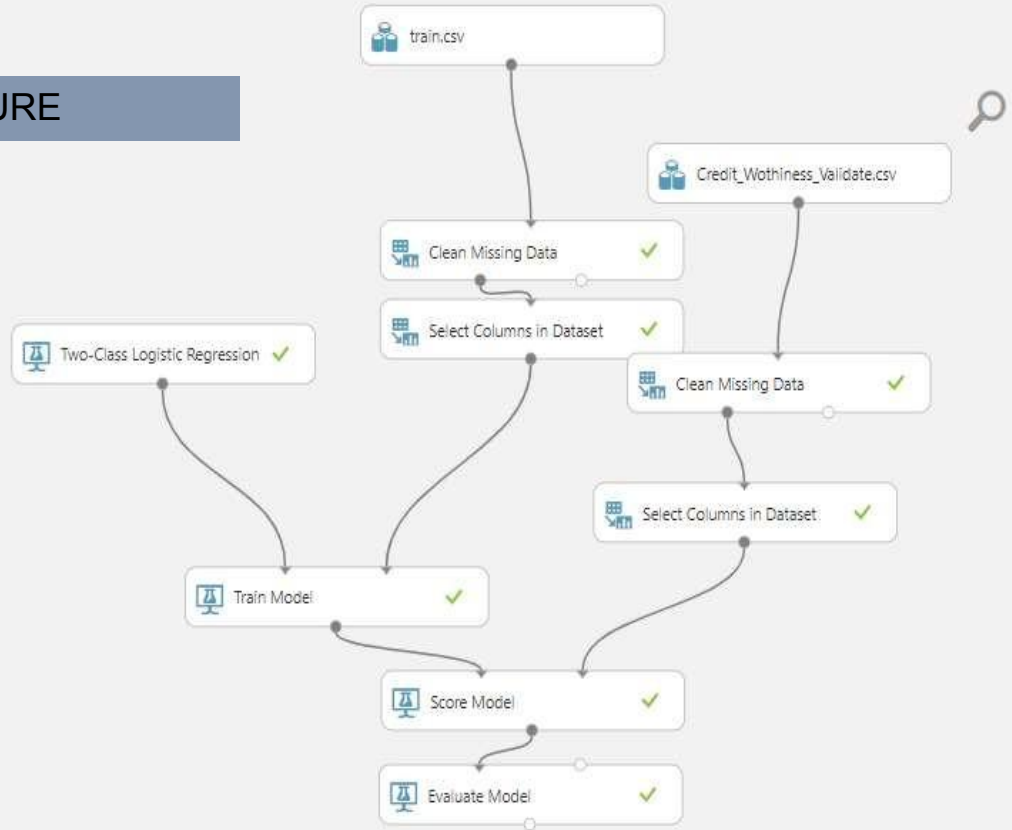
ImplementationDetails(Model)

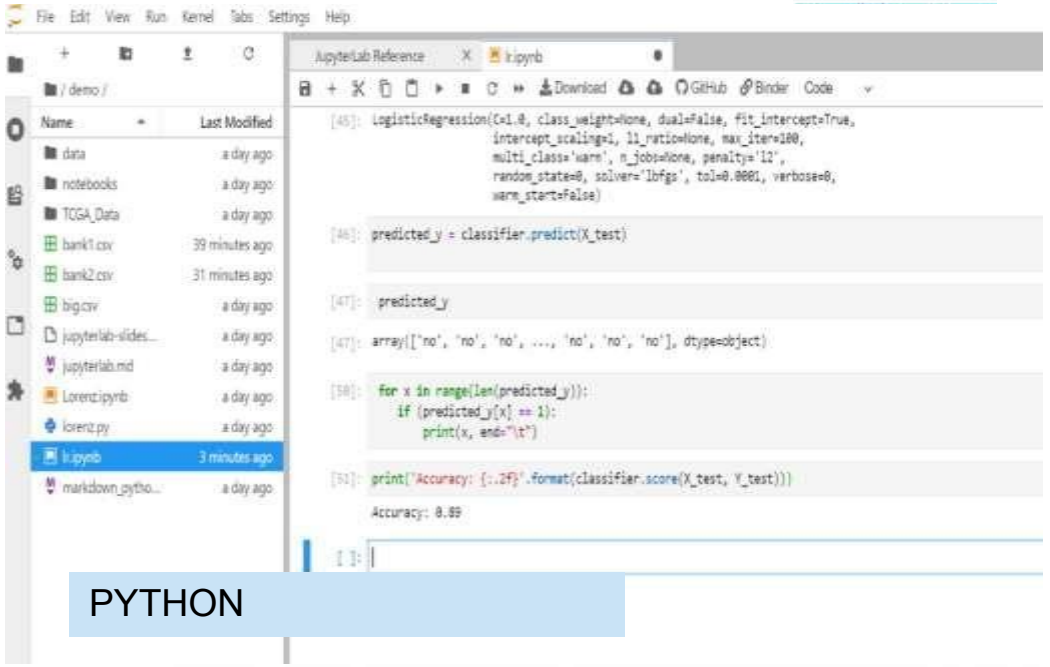
Training experiment | Predictive experiment

Credit worth - C 11

Finished running

MICROSOFT AZURE





PYTHON

LR FORMULA

$$P(x) = \frac{e^{\beta_0 + x \cdot \beta}}{1 + e^{\beta_0 + x \cdot \beta}}$$

We can also re-express the above equation as

$$log(\frac{P(x)}{1 - P(x)}) = \beta_0 + x \cdot \beta$$

IMPLEMENTATION DETAILS

1)Registration/Login/Forgotpswd Page



2) Home/Feature Page

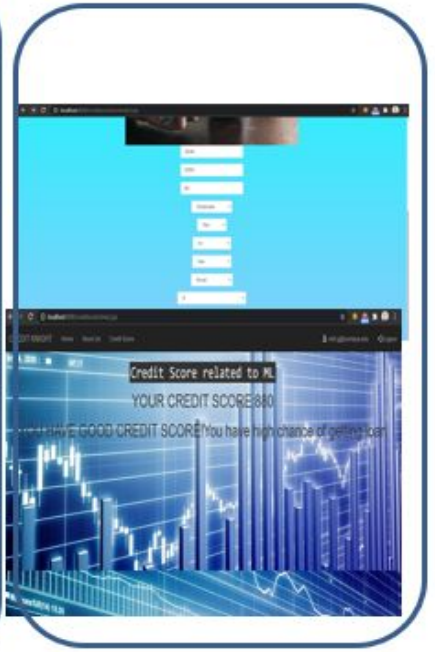
Fico Score

Investment Options

Credit Worthiness

Credit Score using ML

Credit Score Calc.



Results

Algorithm	Accuracy
Logistic Regression	0.94
SVM	0.81
Naïve Bayes	0.80
KNN	0.79
Random Forest	0.78
Decision Tree	0.69

- Table represents the accuracies of different Machine learning Algorithms for Credit Score Calculation and Loan Approval Prediction.
- Logistic Regression has the highest accuracy of 94% in calculating the Credit Score, and thus was used for building the model.
- 12-15 inputs which includes their Gender, Marital Status, Age, Education, Number of Dependents, Applicant Annual Income, Co Applicant Income , Property details etc were used to Calculate the user's Credit Score and Loan Approval request.
- The Credit Score obtained can also be used to find out various investing options user can opt for to improve its credit score and thus increases it chance of loan approval.

TEST CASES

Purpose	Inputs	Expected Output	Procedure
Login into System with correct registered email and password	Registered Email id and Password	Successfully Log in into System	Input Credentials by user are verified with the user data stored in database
Login into system with unregistered Email Id	Unregistered Email id	Account is not registered	Input Email id is checked in the database
Login into system with registered email and incorrect password	Registered Email and Incorrect Password	Wrong Password Entered	Password for the input email is verified with password stored in database for same
Calculation of Credit Score with valid input data	Valid input data in the form	Form Submitted successfully and a output page with the valid credit score.	Input values passed all form validation constraints and data is stored in dataset.

TEST CASES

Purpose	Inputs	Expected Output	Procedure
Credit Worthiness with valid input data	Valid input data in the form	Form Submitted successfully and a output page with the user's credit worthiness	Input values passed all form validation constraints and data is stored in dataset.
Investment Options with a valid Credit Score	Valid Credit Score	Form Submitted successfully and a output page with the all Investment options.	Credit Score is checked whether it lies within the specified range and investment options are available for entered score.
Investment options with invalid Credit Score	Input Credit score out of bound range	Invalid Credit Score	Credit Score entered is checked whether it lies within the specified range.
Loan Eligibility with valid inputs.	Valid input data in the form	Form Submitted successfully and a output page with the valid credit score.	Input data passed all form validation constraints and data is stored in dataset.

Conclusion

- Successfully developed a platform to Calculate User's Credit Score and Predict Loan Approval Request.
- Traditional Methods of banking and financial institution using statistical calculations for identifying customer's financial behavior obtained fraudulent cases which were overcome by our system by calculating the user's Credit Score using various Machine Learning Algorithms.
- An individual's all basic and financial information were considered for calculation of its credit score which gave more accurate results.
- Accuracies of all Machine Learnings Algorithms like Logistic Regression,Naive Bayes,SVM,KNN,Decision tree, Random Forest were obtained and compared.
- Logistic Regression with highest accuracy of 94% was used to develop the model.
- The Credit Score obtained was also used to explore various investing options to improvise the Credit score and thus improve Market Credibility and improve loan approval chances.

References

- Jasmina Nalić and Amar Švraka , **“Using Data Mining Approaches to Build Credit Scoring Model”** ,17th International Symposium INFOTEH- JAHORINA, 21- 23 March 2018.
- Durgesh Kumar Singh and Noopur Goel, **“Analysing Data Mining Techniques on Bank Customers for Credit Score”**, 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), June 2020.
- Archana Gahlaut, Tushar , Prince Kumar Singh , **“Prediction analysis of risky credit using Data mining classification models”,** 8th INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION AND NETWORKING TECHNOLOGIES (ICCCNT) 2017 IIT Delhi , July 3-5, 2017.
- Beibei Niu, Jinzheng Ren , Xiaotao Li , **“Credit Scoring Using Machine Learning by Combing Social Network Information: Evidence from Peer-to-Peer Lending”** ,College of Economics and Management, China Agricultural University, Beijing 100083, China, 17 December 2019.
- Ansen Mathew , **“ Credit Scoring Using Logistic Regression”** ,San Jose State University SJSU ScholarWorks,Master's Projects 532,24 th May 2017.
- Pratik Sharma, Sunil Bhatia, Rohit Burman, Santosh Hazari, Rupali Hande, **“Credit Scoring using Machine Learning Techniques”** , International Journal of Computer Applications (0975 – 8887) Volume 161 – No 11, March 2017.
- ROHIT KUMAR, MAYUKH BISWAS, SANGRAM MONDAL, **“Exploratory Analysis For Credit Score With Applied Machine Learning”**, SRM Institute of Science and Technology, Ramapuram Campus, Chennai,May 2020.

Documents

- Software Project Management Plan
- Software Requirement Specifications
- Software Design Document
- Software Testing Document

Thank You