2023

Research Report on Assessing Customers Applications to Predict Creditworthy and Defaulters

Credit Assessment of each Application to Predict Creditworthy and Defaulters

SHIVANSH SRIVASTAV

Executive Summary

The students of Ivy league University have decided to embark on a start-up journey. They decided to establish and operate a departmental store named Tabz Departmental Store. This store offers a wide variety of products and services that caters to the daily needs of all the people staying in the city in which the university is located. Apart from Tabz, there are a couple of other departmental stores in the city. In order to tackle the intense competition, Tabz decided to offer its products and services on credit i.e., customers can consume products and services without paying for it at the time of consumption. Payment for the products and services consumed from Tabz during the month can be made at the end of the month. This is a revolutionary idea when compared to the traditional mode of cash payment and is expected to simplify the lives of people in the city while giving competitive advantage to Tabz. So, the founders decided to launch a co-branded credit card in association with a local bank, Banco. Tabz has decided to offer two types of cards:

Charge card: The balance is required to be paid in full each month

Lending card: Lending cards allow the customer to pay the balance over a period of time subject to interest being charged

An individual can apply for only one of the two types of credit card on offer. In order to extend the credit card to the individuals, Banco must first underwrite the applicant.
"Underwriting is the process by which the lender decides whether an applicant is creditworthy and should receive a credit line."

Given the innovative business model of Tabz and the sound reputation of its founders, thousands of residents in the city submitted their application forms for the co-branded credit card from Tabz. Along with the data present in application forms, Banco also has access to the consumer bureau. Bureau is an agency that aggregates consumer borrowing and payment information for the purpose of assessing credit-worthiness of an individual and setting a limit on the cumulative credit that can be extended to an individual by lenders. Banco has hired

you to help underwrite each applicant and predict the credit worthiness of an individual. Banco has provided you with the customer application and bureau data with the default tagging i.e., if a customer has missed cumulative of 3 payments across all open trades, his default indicator is 1 else 0. Data consists of independent variables at the time T0 and the actual performance of the individual (Default/ Non-Default) after 12 months i.e., at time T12. Banco's expectation from you is to predict if an applicant will go default in next 12 months from the time of application submission.

Table of Contents

CHAPTER 1 INTRODUCTION, SCOPE AND BACKGROUND

1. INTRODUCTION, SCOPE AND BACKGROUND

1.1 Overview of Project Case / Business case

Main Objective of the Project is to develop a framework that could flag potential defaulters based on rule-based system for bank to take appropriate action from all the application submitted to review.
After reviewing each application and selecting creditworthy application bank will decide whether the applicant is eligible for the Charge Card or Lending Card.

1.2 Problem definition

You have to create a list of applications in the order in which Banco should process them. With an objective to maintain healthy financials, Banco would like to process least risky applications first. Against each application, you also have to provide your prediction of default - 1 or 0, where 1 indicates a default and 0 indicates no default.

Assume:

- A resident of the city can submit only a single application form
- None of the applications submitted are fraudulent

1.3 Project Scope

Nowadays when startups are rising so rapidly in INDIA and worldwide Credit system is a very important and functional function.

In this digital era This is a revolutionary idea when compared to the traditional mode of cash payment and is expected to simplify the lives of people in the city while giving competitive advantage to Tabz. So, the founders decided to launch a co-branded credit card in association with a local bank, Banco.

This framework will help all the startups who are starting their journey or who has established their business to grow as people prefer credit system more than cash mode.

CHAPTER 2
REVIEW OF LITERATURE

2. REVIEW OF LITERATURE

2.1 Literature Review

Literature review for this project was conducted during the spring of 2023. The primary focus was on literature addressing the process, techniques and limitations doing assessment of each application submitted to bank for credit card.

However, literature that generally discusses modelling inputs, including data sources, data collection methods, data parameters are included as long as they are relevant to the credit assessment of each application for Card.

Appendix A identifies literature sources relevant to each parameter and briefly discuss the applicability of the source to developing guidance, data for selected input parameters (input) and tools for the job. Parameters are discussed in the following categories:

- The data dictionary excel sheet will give you the description of all the variables contained in the 3 datasets that have been used in model training and predicting the results.
- Mvar1, Mva2, Mvar3, Mvar4, Mvar5, Mvar6, Mvar7, Mvar8, Mvar9, Mvar10, Mvar11, Mvar12, Mvar13, Mvar14, Mvar15, Mvar16,
- Function and description of these input parameters are defined in an excel sheet for all the 3 databases.
- Some of these parameters have missing and null values as well.
- The data cleaning is done to handle that data after reviewing literature of the project and prepared documentation.

Appendix A (i) also includes a review of some surveys done by the company to better understand the market and the applicant and these surveys contain data crucial for stakeholders and managers to take appropriate action.

Appendix B (ii) includes an annotated Bibliography of the referenced literature sources, with a good general description of each source.

Introduction: Examining consumer's behaviours in choosing and using credits card, this report was commissioned to present an empirical review of Customers willing to apply for credit card and bank wants to assess all the applications for defaulters and creditworthy applications and wants to go with least risky applications first to provide credit facility to them. This literature review on credit card market and applicants with the intention of providing insight for policy holders and implementers in protection within the market.

I reviewed the literature on credit card facilities, consumption, credit history from the perspective of economics, finance, marketing, project research and psychological perspective and discuss (i) how consumers source for and switch between credit card contracts among different finances and contracts, along with (ii) commonly mistakes the determinants of

borrowing behaviour using credit card (iii) how consumers repay credit card debts and thoughts of doing frauds with impact of minimum repay requirement presentation. The extent to which behavioural biases drive consumers' behaviour. Most of this literature review focuses on finding of credit market and bank frauds.

Result and Findings in point: Results and findings of this literature review is totally based on the above-mentioned facts and the study on credit market, bank statements and consumer behaviour study and below are some of the results/findings of those research/review study mentioned:

- Searching and switching consumer action and cost
- Borrowing behaviour
- Credit history legit or fraud
- Repayment option and Behaviour of consumer
- Profit from Borrowing cost
- Behavioural bias and psychological study documentation

Conclusions and Recommendations: With increasing access to personal credit, business credit and credit cards are now pervasively held by most consumers in developing or developed countries to grow their businesses and to increase the plastic money value. This report/review summarizes the academic findings on credit card choices and assessing of defaulters and creditworthy from the applications and use by market. The credit card market is becoming more competitive and consumers are treated very special with discounts and digital points to capture the bigger market. Consumers are generally behaving rather rationally to easy their lives and maximize their utility.

However, consumers are still shown to misuse their credit card, credit history and banks, make silly and obvious mistakes and suffer from behavioural biases so suffering exploitation by the banks. This review draws some specific recommendations that implementors may rely on for their policy making and lower risks to enhance consumer protection. It is important to increase beneficial activity as consumers are searching for low-rate credit cards, cost of repayment as low as possible, and increasing consumer awareness.

2.2 Feasibility Analysis

Feasibility study determines whether it is worth to carry out the intended work. A SWOT analysis can be done for the project. This stage includes identification of business value of the product, evaluation of technical resources/technology needed to solve the existing problem. Under feasibility study an assessment of ethical factors are also considered.

This study is simply an assessment of the practically of a proposed project plan or method. This is done by analysing technical, economic, legal, operational and also time feasibility factors. Just as the name implies, "Is this feasible?

A project feasibility study should be done during the project management life-cycle after the business case has been completed. An effective feasibility study points a project in the right direction by helping decision makers have a holistic view of potential benefits, disadvantages also and, constraints that could affect it's outcome. This study is important to decide whther the project can be not only viable but also beneficial from a technical, financial, legal market standpoint.

A feasibility Study reports following elements:

i. Executive Summary
ii. Description of Product
iii. Technology and Tools consideration
iv. Project Planning and Designing
v. Implementing Research methodologies
vi. Results and Findings

Technical Feasibility: Technical feasibility is an assessment of whether a proposed project, product, or service can be successfully implemented using current or available technology. It involves evaluating in detail the technical requirements, constraints, and capabilities of the proposed solution/process to determine whether it is feasible to develop, implement, and maintain it within the given constraints and available resources.

1. **Identifying potential problems:** This report can help identify potential problems and challenges that may arise during the implementation project. This can help project managers in planning for contingencies and mitigate risks before they become major issues.
2. **Assessing resource requirements:** A technical feasibility study helps assess the resources required for a project, such as equipment, software, raw materials and personnel. This can help project managers estimate costs, plan budgets, and allocate resources more effectively.
3. **Evaluating technical requirements:** This helps evaluate the technical requirements of a project, such as hardware and software compatibility, data management, production process and security. This can help ensure that the project is technically sound and can be successfully implemented.

4. **Ensuring project viability:** It helps determine whether a project is viable in terms of its technical feasibility. This can help project managers decide whether to proceed with the project or consider other alternatives.

The essential questions that help in testing the operational feasibility of a system include the following:

- Is the project feasible within the limits of current technology?
- Does the technology exist at all?
- Is it available within given resource constraints?
- Is it a practical proposition?
- Are the current technical resources sufficient for the new system?
- Do we possess the necessary technical expertise, and is the schedule reasonable?
- Can the technology be easily applied to current problems?
- Does the technology have the capacity to handle the solution?
- Do we currently possess the necessary technology?

Overall, technical feasibility is essential for any project because it helps identify potential problems, assess resource requirements, evaluate technical requirements, and ensure project viability. By conducting this study, project managers can improve the chances of project success and avoid costly mistakes.

Cost Benefit Analysis: The way that many businesses, organizations, and entrepreneurs answer whether a particular decision is the best one for your business? Will it make success? and other, questions are through business analytics—specifically, by conducting a cost-benefit analysis. A cost-benefit analysis is the process of comparing the projected or estimated costs and benefits (or opportunities) associated with a project decision to determine whether it makes sense from a business perspective.

Generally speaking, cost-benefit analysis involves tallying up all costs of a project or decision and subtracting that amount from the total projected benefits of the project or decision.

If the projected benefits outweigh the costs, you could argue that the decision is a good one to make. If, on the other hand, the costs outweigh the benefits, then a company may want to rethink the decision or project.

Operational Feasibility: Operational feasibility is a measure of how well a proposed system solves the problems, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development.
Operational feasibility reviews the willingness of the organization to support the proposed system. This is probably the most difficult of the feasibilities to gauge. In order to determine

this feasibility, it is important to understand the management commitment to the proposed project

The essential questions that help in testing the operational feasibility of a system include the following:

- Does current mode of operation provide adequate throughput and response time?
- Does current mode provide end users and managers with timely, pertinent, accurate and useful formatted information?
- Does current mode of operation offer effective controls to protect against fraud and to guarantee accuracy and security of data and information?
- Does current mode of operation make maximum use of available resources, including people, time, and flow of forms?
- Does current mode of operation provide reliable services
- Are the services flexible and expandable?
- Are the current work practices and procedures adequate to support the new system?
- Government regulations
- Are the users not happy with current business practices?
- Will it reduce the time (operation) considerably?
- Have the users been involved in the planning and development of the project?
- Does the overall response increase?
- Will accessibility of information be lost?
- Will the system affect the customers in considerable way?
- Legal aspects

These Feasibility analysis topics will be inserted into the project report but since tis project is not an Enterprise level Implemented software but only a research project. So not many of above-mentioned points will be used in this Project.

CHAPTER 3 PROJECT PLANNING AND METHODOLOGY

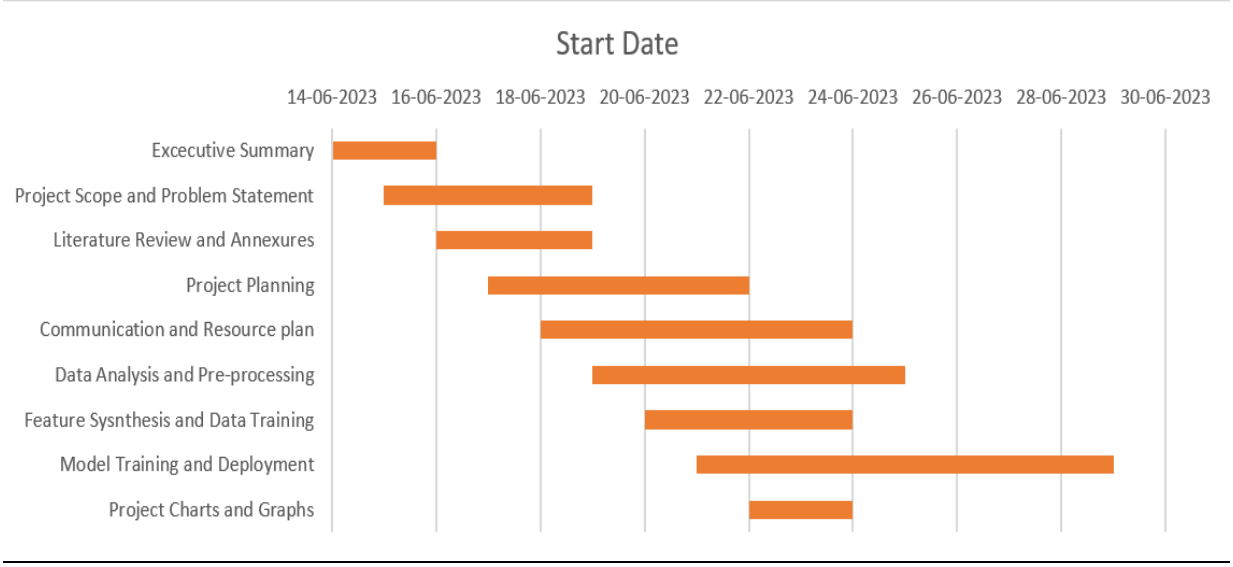
3. PROJECT PLANNING AND METHODOLOGY

3.1 Project Planning

Project Planning is the first step of developing or researching any project. This section is responsible for developing your project in a given time and with selected resources. In this section the following this are being followed and implemented –

- Timetable to complete each task with defining and implementing those within the time frame
- Designing a Gantt Chart for the same. (Attached below)
- Making a Communication Plan to set path for the tasks and to seek time and guidance following that communication plan (Attached below)
- Review the Project plan and Resource Plan for better optimization of tasks and resources to complete the Project.
- Limitations of the Project will be handled at the research phase by time to time as project is being completed.

Gantt Chart of Project Plan



This Gantt Chart Show each Task with it's Start date and End date

Communication Plan of the Project

Communication	Purpose	Audience	Owner	Method Of Communication
Project Selection Meeting	Deciding project for each member of the group so no overlapping is done and to Initiate the Project by deciding Project Title	Members of Group P31 [the group assigned]	Shivansh Srivastav	Zoom Meeting
Mentor Interaction for Approval	Interacting with mentor to submit a report containing all the details about Project title, Problem statement and Project scope to get his approval	Prof. Yerriswamy M, Members of P31	Prof. Yerriswamy M	Zoom Meeting
Project Planning Meeting	Planning the phases of project completion and time allocation to each and every task. Modifications were done afterwards and final schedule is submitted as Gantt chart.	Shivansh Srivastav [Myself]	Shivansh Srivastav	Email and Charts
Team meeting for Project Status	Holding a team meeting for everyone's project status and what help can others provide for their research without taking their work	Selected Members of group P31	Shivansh Srivastav	Zoom Meeting
Project daily Report	Preparing daily logs of project status and what resources and limitations were there to follow.	Shivansh Srivastav [Myself]	Shivansh Srivastav	Email
Review Board Meeting	Review meeting was done 4 weeks from project start date to review the project and whether the research methodologies are	Selected Members of group P31 and Myself	Shivansh Srivastav, P31	Zoom Meeting, Emails

	enough to provide a better result/findings.			
Team meeting to increase Resources	Holding a team meeting to help each other gathering resources and proofs to support the project and its outcomes.	Group P31	Group P31	Zoom Meeting
Mentor Interaction for Project Report	After completing the research from my end Interacting with Mentor to understand preparing a project report and submit for final year project.	Prof. Yerriswamy M, Members of P31	Prof. Yerriswamy M	Zoom Meeting

3.2 Methodology

3.2.1 Qualitative Methods Used:

- Qualitative Research is a method that collects data using conversational methods in most cases.
- Not very useful for Model training as response collected is non-numeric but helps us understand the goal and way of thinking of defaulters.
- This method not only helps a researcher understand what participants think but also their way of thinking as why they think in a particular way.
- Useful in understanding the thoughts (or way of thinking) of defaulters.

i. **Text analysis:** Text analysis is different from other qualitative methods as it is used to analyse constructs decoding from any text through available research documentations. I (The researcher) study and understands the contexts in which the documents are written and then tries to draw meanings and charts to provide resources for the data collection and data feeding.

ii. **Case Study:** Case Study research is used to study an organization or an entity. This method is one of the most valuable options for modern this type of research is used in education sector, psychological sector, financial sector and, philosophical sector as well.

This method provides a very deep dive into collecting data and in ongoing research as well.

I feel that a single estimator is prone to overfit the data and hence might not perform well on the test data.

Hence, I used ensemble (averaging over different estimators) of models that are different in nature to reduce overfitting and give us a robust estimator as a method of researching the project.

3.2.2 Quantitative Methods Used:

- Quantitative methods deal with numbers and with measurable forms of data.
- It uses a very systematic way of researching data and events.
- Quantitative methods are used to answer questions in terms of justifying relationships with measurable variables to either predict, explain or control a phenomenon.

i. **Survey Research:** The ultimate and very goal of survey research is to learn about number of populations who wants or do not want the credit card that is being issued by the bank banco. Nowadays online surveys are in trend via email and social media platforms but any mode of survey works just fine to gather data.
In this method the researcher designed most relevant questions to the product and collects the survey to take a decision and whether to launch the product and done actual data collection and prediction of defaulters and creditworthy applications.

ii. **Descriptive Research:** This is a method which identifies the characteristics of an observed phenomenon and collects more information on the research area. In simple words, descriptive research is all about describing the phenomenon, observing it, and drawing conclusions from it.

iii. **Correlation Research:** Correlation research examines the relationship between two or more variable. There are two types of correlation one is positive correlation that has a relation with the variable and other is negative correlation that has no direct relation to the variable. This will help in understanding us the correlation between the parameters of our research.

CHAPTER 4
DATA ANALYSIS, PREPROCESSING AND FEATURE SYNTHESIS

4. DATA ANALYSIS, PREPROCESSING AND FEATURE SYNTHESIS

4.1 Requirement Analysis

4.1.1 Data Collection

Data collection is the main and crucial part of any project and as the saying is **'Feed in garbage and you will get garbage'.** So, collecting data that has some relation with the topic is very important for the project completion so that you can get better results and findings. Otherwise, it will be very tough to find the optimal solution of the research.
I have collected data from different sources and these sources are mentioned below:

- American Express open research financial datasets
- Kaggle
- Through third party APIs so no illegal mean is used to collect the data

4.1.2 Data Analysis and tools of data analysis

Following files were downloaded for analysis of the topic and research to find results:

Training_dataset.csv: This dataset contains:

- Applicant level historic credit history
- Performance in terms of default tagging i.e., 1 for default and 0 for no default
- Application and bureau data

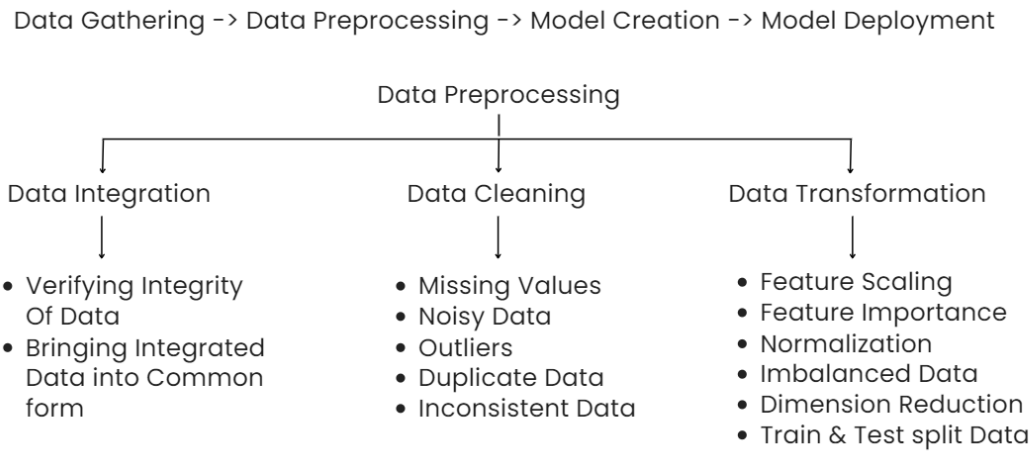
Leaderboard_dataset.csv: This data has historical applicant level data along with all the variables in the training dataset. The actual performance i.e., default tagging is not present in this data.

Evaluation_dataset.csv: This data has applicant level data along with all the variables in the training dataset. The actual performance i.e., default tagging is not present in this data.

Data_Dictionary.xlsx: This sheet will give you the description of all the variables contained in the 3 datasets that have been used in model training and predicting the results.

4.2 Data Pre-processing

Used transformation (log or cube root) to make the distribution Gaussian as well as normalized 21 feature columns including the ones which I got after removing and combining some of the input data – explained with below image:



i.Img1

4.3 Estimation Techniques Used

I used a Gradient Boosting Machine – implementing the XGboost Python library.

A Gradient Boosting Machine iteratively trains decision trees, and ensembles them (essentially combining their predictions), all while minimizing the error function (in my case, the 'softmax' function) at each step. The XGboost library provides an out-of-the-box implementation of a GBM. It is a form of extreme Gradient Boosting.

We then used an ensemble model of Linear Support Vector Machine, Extreme Gradient Boost, Multi-Layer Perceptron and Random Forest. Each of these models are alone powerful models and combining them increases our predictive power.

Ensemble methods helps improve machine learning results by combining multiple models. Using ensemble methods allows to produce better predictions compared to a single model.

Reasons for using these Techniques

I feel that a single estimator is prone to overfit the data and hence might not perform well on the test data.

Hence, we used ensemble (averaging over different estimators) of models that are different in nature to reduce overfitting and give us a robust estimator.

4.4 Data Visualization

Feature Importance using Random Forest Feature Sampling Technique

Correlation Plots between Features

4.5 Feature Synthesis [Detail of each variable used in the logic/strategy/model]
Nearly all the variables had missing values. I analysed all the variables and decided that the mean of the features was the best alternative for the missing values.

The variables mvar6 was found to be inversely related to the default rate whereas mvar7 and mvar8 were directly related to it. So, I used a new feature (mvar7 + mvar8)/mvar6.

Similar observations led me to include new features like (1+mvar3) * (1+mvar4) * (1+mvar5) and (mvar36) * mvar(37)/mvar38. These synthesised features were helpful in increasing the performance which can be seen from the feature importance graph in above section.

To reduce the number features I removed highly correlated features using the heatmap from correlation plot (on above section). For e.g.: we removed mvar16 and mvar17 as they were highly correlated with mvar18. These reductions made our model faster and further increased the performance.

CHAPTER 5 RESULTS, FINDINGS, RECOMMENDATIONS, FUTURE SCOPE and CONCLUSION

5. RESULTS, FINDINGS, RECOMMENDATIONS, FUTURE SCOPE and CONCLUSION

5.1 Results of the project work

Overall project evaluation is the data collected through the methods were very inconsistent and noisy and after applying many tools and methods the data we trained for model prediction predicted a significant amount of Defaulters and Credit worthy applications.

This research project is very useful in tracking and finding the customers with better credit history so that a company does not have to bear a loss.

5.2 Findings based on analysis of data

In this section, findings of the projects are mentioned and those findings are-

- More than 26 % of the applications were fraud.
- A significant amount of people applied for both Charge and Lending card.
- This Project predicted the number of both frauds and Creditworthy applications with a maximum of 97.1%

Models used in my project to predict better results:

I have used most of the machine learning and deep learning models so that I can get a better idea of which model gives the better result and which optimizes the solution more. Below mentioned models are used in my project with asymmetric order -

- Random Forest
- Gradient Boosting Machine
- XGBoost - from CoreLib and sklearn API [They do give different accuracies so better to try them both]
- SVM Regressor
- Categorical Variable Analysis
- Data Exploration
- Feature Synthesis
- Data Normalization
- Ensembling of the best models
- K-fold Cross-Validation

5.3 Recommendation based on findings

I would recommend this project to any company whether that is low-level company, a middle-level company or a multi-national company that wants to provide a method different from cash mode such as credit system because it will give you the 99% success rate of finding and distinguishing between frauds and creditworthy applicants.

5.4 Suggestions for areas of improvement

The only suggestion that came to my mind while researching and developing this project was that banks and other financial companies who grants credit cards need to improve their method of selecting creditworthy applicants and set the bar to a more suitable level.

5.5 Scope for future work

In this section, I am mentioning scope of this project in future as Credit system is going to increase in time so will the frauds. So, this project will provide a better tracking system for checking application credibility.

5.6 Conclusion

From this Project I have concluded that startups and companies spend a lot of time and resources while giving out Credit cards to grow their business as there were significant amount of number of fraud applications looking to get cards with bad credit or applying for both the cards at same time.
My Conclusion is a better and strong financial software should be used to track and grant the card to only the creditworthy.

BIBLIOGRAPHY

American Express analyse the problem statement (2017), Kaggle

Surveys, Qualitative and Quantitative research references, Professor and team

Reports, Articles on credit frauds and card benefits, Manuscript produced/developed by the researcher of the Project dataset, Bureau of financial advisory, Intelligence Bureau

Technical Feasibility, Economic Feasibility, Operational feasibility, Legal Feasibility Retrieved August 13, 2011 from http:www.freetutes.com/systemanalysis/sa3- technical-economic-operational-legal.html

Compiled By

Shivansh Srivastav