INTERNSHIP AND ITS IMPORTANCE

An internship is a platform offered by an employer to potential employees to work at a firm for a fixed, limited period of time. The employees who work for internship are called interns. Interns are usually undergraduates or students. Most of the internships last for any length of time between one week and 12 months. Internships may be part-time or full-time. Internships offer practical experience for any student in the industry relating to their field of study. This experience is valuable to students allowing them to experience how their studies are applied in the "real world".

An intern is one who works in a temporary position for an employer that operates in an industry they are interested to work. Unlike any conventional employment, internships have an emphasis on training, rather than employment itself. An internship provides opportunity for employees to acquire experience in a particular field or in any industry, determine if they have an interest in a particular area of interest. Interns may also have the possibility of putting themselves forward for upcoming opportunities for paid work, during their internship.

Companies offer students internships for a variety of both short and long-term reasons. In the short-term, internships provide employers with small works like office based tasks, such as photocopying, filing or report drafting. Long-term, employers can use internships as an effective way of advertising their graduate jobs and/or schemes to students. An internship may be paid, unpaid or partially paid. Paid internships are most common at engineering, legal, business like accounting and finance, technology, medical, science, and advertising sectors. Internships in the media like mass media, print media and non-profit organizations are often unpaid.

A research internship is also known as a dissertation internship. It is usually undertaken by a student in his last year of academic study. For a research internship, a student will undertake research for a particular company.Internship opportunities in India are career specific. College students often choose internships based on their branch of study. This is a unique learning opportunity that you may never have again as a working adult. Lastly one can gain confidence in his own abilities.

CHAPTER 1

OVERVIEW OF THE ORGANIZATION

Company Name: Excelsoft Technologies Pvt Ltd.

About the Company:

Excelsoft Technologies is a provider of innovative technology-based solutions in the education and training space. It architects, designs and develops technology solutions and digital content and has established itself in a leadership position in the e-learning business. The company has made extraordinary contributions towards effectively using technology to enhance capacities of teachers and learners in India and the world over.

Incorporated in the year 2000, Excelsoft designs E-learning solutions based on a pedagogical approach towards content design and technology implementations. It caters to the education and training management requirements for diverse industry base of School Education, Higher Education, Vocational Education, Corporate, Government, Defense, Educational Publishers and many more in Test and Assessment delivery.

Excelsoft adopts an innovation-centric approach towards solution design. With deep domain expertise it provides a solution-oriented approach and hence does not offer canned products like most competitors. This helps in effectively addressing the pain points and challenges faced by the end-client and provide the agility and flexibility to architect the solution.

Excelsoft has a best fit solution approach, by combining more than one product for a particular solution, unlike competing brands. Adoption of newer technology trends well-ahead of the completion has helped Excelsoft deliver solutions that are future-ready and hence can be enhanced and scaled-up to suit the requirement.

With operations in India, Malaysia, Singapore, UK and the USA, Excelsoft is continuously increasing presence of both client base and operational offices across the globe. The company is looking at continued investment in new product developments and increased services. It is also hopeful about strategic acquisitions of companies operating in the E-learning domain.

Company Started Year: 2000

1.1 SERVICES

- E-Learning
- Authoring Tools
- Saras - LMS, AMS, PMS, TMS, Test and Assessment Solutions
- School Solution
- Curriculum development
- Digital Books
- Training Management
- Cognowise - Learning Analytics
- Consulting,
- Mobile Learning & Content Solutions
- Augmented reality, virtual reality,
- Adaptive learning, eBooks,
- K12 education and Higher Education

CHAPTER 2
TRAINING PROGRAM

The Company trained us for four weeks in which we were taught the basics of Python programming and helped us gain knowledge on the fundamentals of Machine Learning.

1. Python: The basic concepts of Python were revised and different Python libraries and their applications were discussed.

2. Machine Learning Fundamentals: Definition of Machine Learning, its existence, uses, applications and types of Machine Learning Techniques were discussed.

3. Supervised Learning: The company concentrated more on supervised learning as the domain for the training program and helped us work on a project involving supervised learning.

4. Datasets and Workflow: The company taught us the importance of data and dataset preprocessing before actually applying the algorithms on it. They also taught us how to approach a machine learning problem methodically and logically.

5. Scikit-learn: Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use. The company taught us to use this library for our project.

2.1 PROJECT OVERVIEW

The purpose of *Loan Prediction using Machine Learning* is to develop a quick, immediate and easy to use technological tool to choose deserving applicants for loans from banks. The Loan Prediction System can automatically calculate the weight of each features taking part in loan processing and on new test data same features are processed with respect to their associated weight A time limit can be set for the applicant to check whether his/her loan can be sanctioned or not.

Loan Prediction System allows us to verify specific applications so that they can be checked on priority basis. This project is exclusively for providing loan for people based on different criteria and the whole process of prediction is done privately, so that no stakeholders would be able to alter the processing. Results against particular Loan Id can be sent to various department of banks so that they can take appropriate actions on application. This helps all other departments to carry out other formalities.

2.2 OBJECTIVES

- Ensuring data integrity and process integrity.
- Automate the whole process of verification and validation of a loan applicant.
- Predict whether a particular applicant is a safe play and if he/she deserves the loan.
- Usage of Machine Learning Algorithms to predict the credibility of an applicant.

CHAPTER 3
INTRODUCTION

3.1 MACHINE LEARNING

Machine Learning (ML) is that field of computer science with the help of which computer systems can provide sense to data in much the same way as human beings do. In simple words, ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method. The main focus of ML is to allow computer systems learn from experience without being explicitly programmed or human intervention.

The formal definition of Machine Learning as given by Prof. Tom Mitchell is "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

The above definition is basically focusing on three parameters, also the main components of any learning algorithm, namely Task (T), Performance (P) and experience (E). In this context, we can simplify this definition as −

ML is a field of AI consisting of learning algorithms that −

- Improve their performance (P)
- At executing some task (T)
- Over time with experience (E)

3.2 SUPERVISED LEARNING

Supervised machine learning algorithms are designed to learn by example. The name supervised learning originates from the idea that training this type of algorithm is like having a teacher supervise the whole process. When training a supervised learning algorithm, the training data will consist of inputs paired with the correct outputs. During training, the algorithm will search for patterns in the data that correlate with the desired outputs. After training, a supervised learning algorithm will take in new unseen inputs and will determine which label the new inputs will be classified as based on prior training data. The objective of a supervised learning model is to predict

the correct label for newly presented input data. At its most basic form, a supervised learning algorithm can be written simply as:

$$Y=f(x)$$

Where *Y* is the predicted output that is determined by a mapping function that assigns a class to an input value *x*. The function used to connect input features to a predicted output is created by the machine learning model during training. Supervised learning can be split into two subcategories: **Classification** and **Regression**.

3.3 K- NEAREST NEIGHBORS ALGORITHM

KNN is a non-parametric, lazy learning algorithm. Its purpose is to use a database in which the data points are separated into several classes to predict the classification of a new sample point. KNN Algorithm is based on feature similarity: How closely out-of-sample features resemble our training set determines how we classify a given data point. KNN can be used for classification — the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. It can also be used for regression — output is the value for the object (predicts continuous values). This value is the average (or median) of the values of its k nearest neighbors.

3.4 PYTHON FOR MACHINE LEARNING

Python is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code. Python is a programming language that lets you work quickly and integrate systems more efficiently.

Python is the fifth most important language as well as most popular language for Machine learning and data science. The following are the features of Python that makes it the preferred choice of language for data science −

- **Extensive set of packages**

  Python has an extensive and powerful set of packages which are ready to be used in various domains. It also has packages like **numpy, scipy, pandas, scikit-learn** etc. which are required for machine learning and data science.

- **Easy prototyping**

  Another important feature of Python that makes it the choice of language for data science is the easy and fast prototyping. This feature is useful for developing new algorithm.

- **Collaboration feature**

  The field of data science basically needs good collaboration and Python provides many useful tools that make this extremely easy.

- **One language for many domains**

  A typical data science project includes various domains like data extraction, data manipulation, data analysis, feature extraction, modelling, evaluation, deployment and updating the solution. As Python is a multi-purpose language, it allows the data scientist to address all these domains from a common platform.

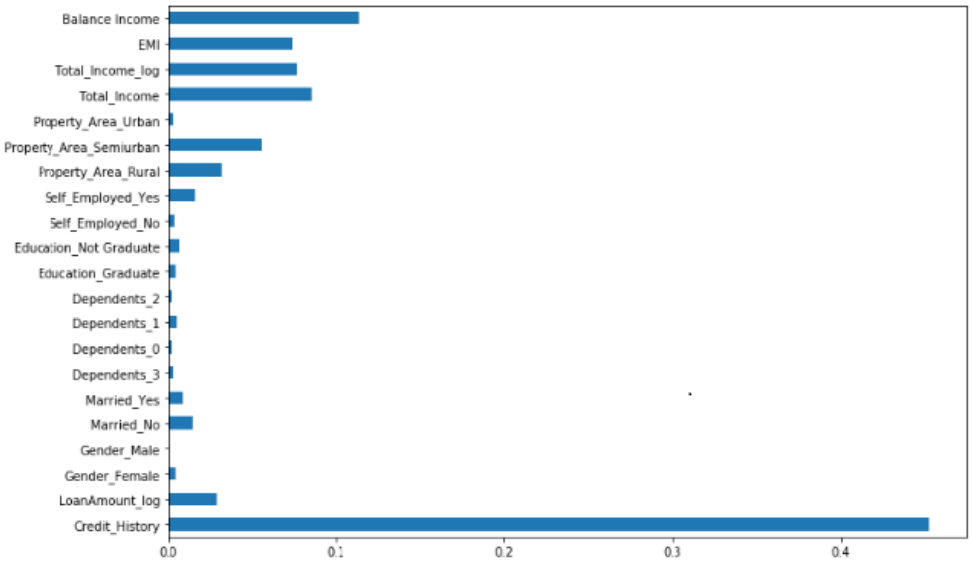
3.5 SCIKIT LEARN

Scikit-learn is a robust library used to bring Machine Learning with Python into a production system. Scikit-learn was initially developed by David Cournapeau as a Google summer of code project in 2007. Later, Matthieu Brucher joined the project and started to use it as a part of his thesis work. In 2010 INRIA got involved and the first public release (v0.1 beta) was published in late January 2010. The project now has more than 30 active contributors and has had paid sponsorship from INRIA, Google, Tinyclues and the Python Software Foundation.The library is built upon the SciPy (Scientific Python) that must be installed before you can use scikit-learn. This stack includes **NumPy, SciPy, Matplotlib, IPython, SymPy** and **Pandas**. Some popular group of models provided by scikit-learn include: Clustering, Cross Validation, Datasets, Dimensionality Reduction, Ensemble Methods, Feature Selection and Extraction, Parameter Tuning, Manifold Learning, Supervised Models.

CHAPTER 4

SNAP-SHOTS

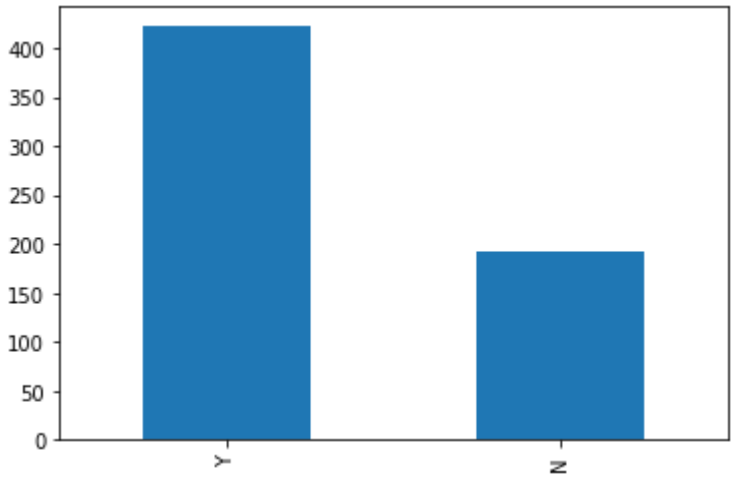
- **Feature Column Graph.**



Snapshot 4.1: Feature Important Columns

Description: Gives a bar plot for easy understanding and feature selection

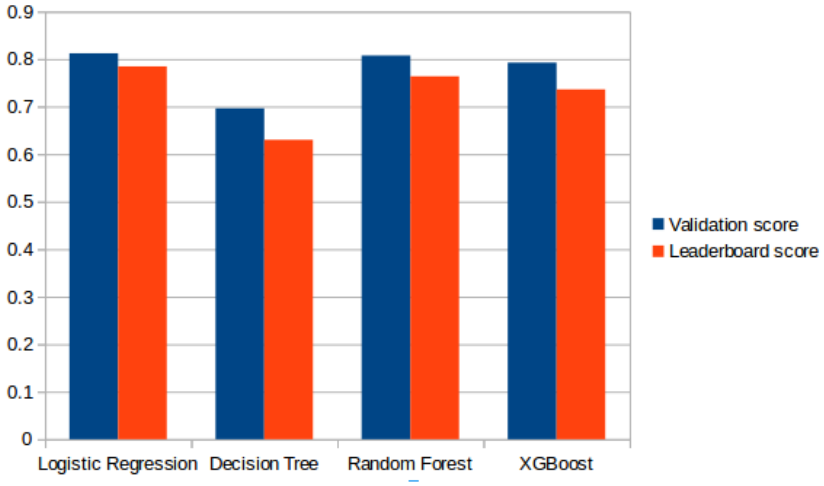
- **Loan Status Graph for Training Data**



Snapshot 4.2 : Loan Status with ‘yes’ or ‘no’ categorical value

Description: This graph is used to identify the number of values of the target variable as well as how many positive and negative examples are there in the training data. Around 422 out of 614 applications have been approved according to the above graph.

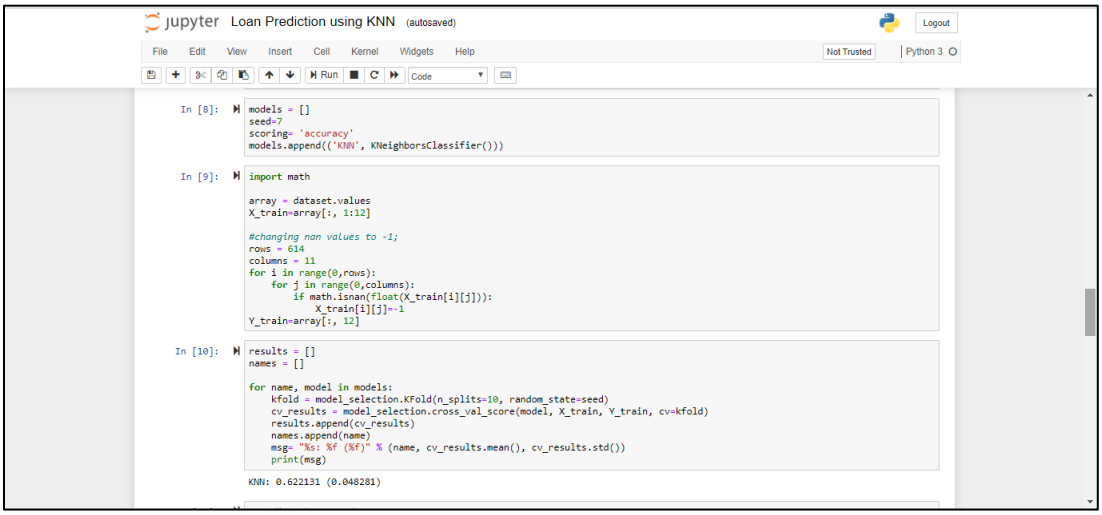
- **Results of various algorithms**



Snapshot 4.3 : Validation and Leaderboard Score

Description: This graph summarizes the validation and leaderboard scores for 4 different algorithms other than KNN. As we can infer from the graph, Logistic Regression achieves the best accuracy (0.7847) followed by Random Forest (0.7638).

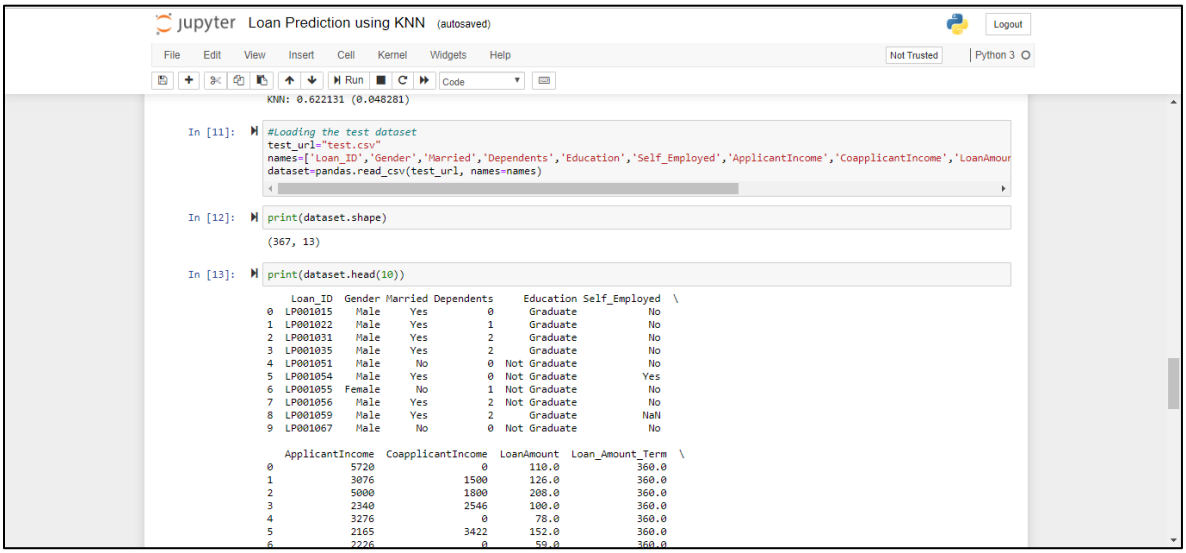
- **Training the model using KNN Algorithm**



Snapshot 4.4: Training the model

Description: This image shows the code used to train the model on the given csv file for loan prediction using K-Nearest Neighbors Algorithm.

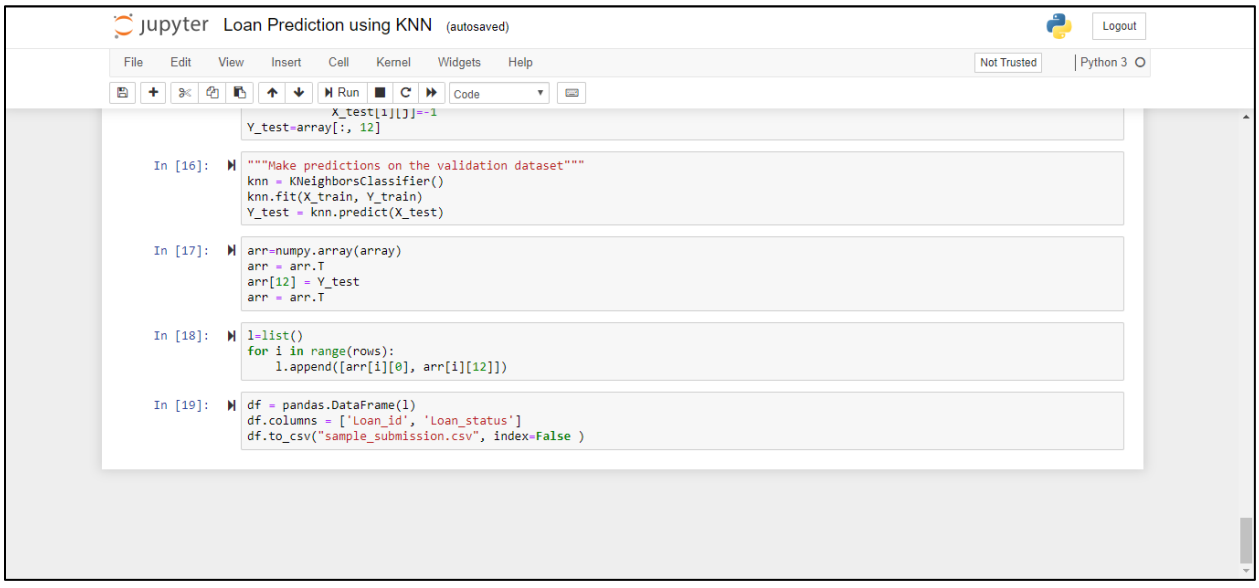
- **Loading the test data for prediction**



Snapshot 4.5: Loading the test data for prediction

Description: This image consists of the code used to import the test data to make predictions and it also displays few important information about the test data.

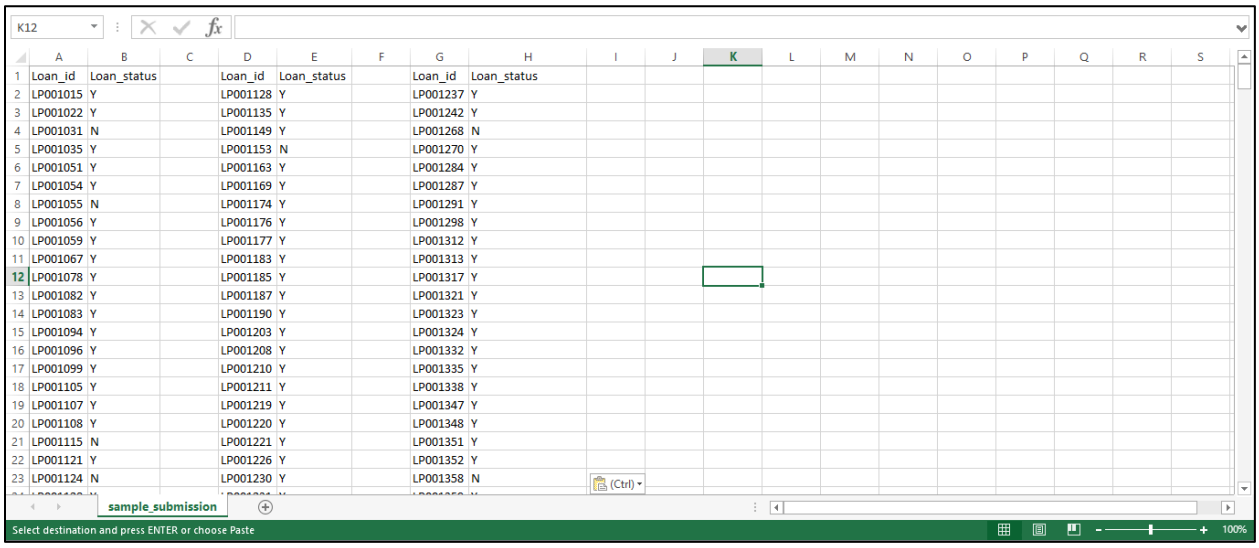
- **Predicting using KNN algorithm on test data**



Snapshot 4.6 Predicting using KNN

Description: This image consists of the code used to predict on the test data using the trained model.

- **Our result after using KNN algorithm for prediction**



Loan_id	Loan_status	Loan_id	Loan_status	Loan_id	Loan_status
LP001015	Y	LP001128	Y	LP001237	Y
LP001022	Y	LP001135	Y	LP001242	Y
LP001031	N	LP001149	Y	LP001268	N
LP001035	Y	LP001153	N	LP001270	Y
LP001051	Y	LP001163	Y	LP001284	Y
LP001054	Y	LP001169	Y	LP001287	Y
LP001055	N	LP001174	Y	LP001291	Y
LP001056	Y	LP001176	Y	LP001298	Y
LP001059	Y	LP001177	Y	LP001312	Y
LP001067	Y	LP001183	Y	LP001313	Y
LP001078	Y	LP001185	Y	LP001317	Y
LP001082	Y	LP001187	Y	LP001321	Y
LP001083	Y	LP001190	Y	LP001323	Y
LP001094	Y	LP001203	Y	LP001324	Y
LP001096	Y	LP001208	Y	LP001332	Y
LP001099	Y	LP001210	Y	LP001335	Y
LP001105	Y	LP001211	Y	LP001338	Y
LP001107	Y	LP001219	Y	LP001347	Y
LP001108	Y	LP001220	Y	LP001348	Y
LP001115	N	LP001221	Y	LP001351	Y
LP001121	Y	LP001226	Y	LP001352	Y
LP001124	N	LP001230	Y	LP001358	N

Snapshot 4.7 Results CSV file

Description: This file consists of the predictions of our KNN algorithm on the test data. Results are stored in the form a CSV file containing Loan_id and Loan_status as columns. To show the results of more number of test records, we have used 3 columns to display the results.

CHAPTER 5
LEARNING EXPERIENCES

5.1 KNOWLEDGE AND SKILLS ACQUIRED

- Identifying problems in the real world and trying to find a solution to them in a systematic manner.
- Fundamentals of Machine Learning and usage of Python Libraries
- Dataset collection and Preprocessing.
- Overview of Supervised Learning and Scikit-learn.

5.2 OBSERVED ATTITUDES AND GAINED VALUES

- Team Work
- Time Management
- Communication skills.
- Commitment towards work.
- Self-confidence.
- Learning with fun.

5.3 THE MOST CHALLENGING TASK PERFORMED

Machine learning was a new domain to learn and hence gaining sufficient knowledge and applying them to the project in order to complete it in the given time was challenging. Working under trained professionals and getting to understand their level of work proficiency was quite challenging too.

CONCLUSION

It was a great experience of being able to work as an intern for this organisation. As team we successfully completed the project that was assigned by the company's professionals. The technical aspects of the work done are not flawless and could be improved provided enough time. Being someone with no prior experience with Scikit-learn our time spent in research and discovering it was well worth it and contributed to finding an acceptable solution to build a well-trained machine learning model for Loan Prediction.

We got an idea about the work ethics and technical skills one should possess before entering an IT industry. We also understood the rules and responsibility one should follow as an IT professional in the future. It was indeed a great exposure to the corporate world.

REFERENCES

[1] https://www.towardsdatascience.com /a-brief-introduction-to-supervised-learning-54a3e3932590

[2] https://machinelearningmastery.com/a-gentle-introduction-to-scikit-learn-a-python-machine-learning-library/

[3] https://blog.usejournal.com/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7

[4] https://www.geeksforgeeks.org

[5] https://www.tutorialspoint.com/machine_learning_with_python/index.html

[6] http://www.excelsoftcorp.com

[7] https://www.linkedin.com/company/excelsoft-technologies/about/