

Predicting Profit of a Startup Companies using Machine Learning Algorithms

Dr. SK. Mahaboob Basha¹,
Professor,
NRI Institute of Technology,
A.P, India-521212

Ch. Sirisha²,
UG Scholar,
Dept. of IT,
NRI Institute of Technology,
A.P, India-521212

R. Manikanta³,
UG Scholar,
Dept. of IT,
NRI Institute of Technology,
A.P, India-521212

P. Ramyasri⁴,
UG Scholar,
Dept. of IT,
NRI Institute of Technology,
A.P, India-521212

Abstract:- The profit earned by a company for a specific period depends on several factors like how plenty of time and money a company spends on research and development marketing and alive more. So forecast the profit of a company for a specific period we need to alternation a machine learning model with a dataset that contains factual data about the profit generated by the company. We can set an achievable goal predicting profit is an important task for every business. For example, if the business allots \$500 on marketing, it can't take a profit of \$20,000. correspondingly, there are lot of other components on which the profit of a business depends. A company should set a goal that can be accomplished. In the section below, I will walk you buttoned up the task of profit prediction with machine learning using Python. When there is no automated system there is always a hard in predicting profit as there are more considerations to be done. Wisdom of Models is a machine learning algorithm for predicting profit via research and development costs, administration cost and marketing spend in a business. Wisdom of Models focusses on other three machine learning algorithms, svr, random forest and linear regression to derive a new prediction. Many companies, institutions, governments, and Private organisations are funding these businesses and enticing people to utilise them to investigate their ideas.

Keywords: *Research and Development, Wisdom of Models, Administration Cost, Marketing Spend, SVR, Random Forest, Linear Regression.*

I. INTRODUCTION

The project titled 'PROFIT PREDICTION' it can predicate the profit. The main aim of the project is to predict the profit. A company should always set a goal that must be achievable, else, for their potential employees will not be able to work if they find that the goal set by the company is unachievable. profit prediction for a particular period we can set a goal of the profit prediction. If you know how much profit you can make with the amount of research and

development and marketing you do, then a business can make more than the predicted profit provided the predicted value is achievable. Task of profit forecast with machine learning using Python. The dataset of profit prediction includes data about the research and development spend, Administration cost, Marketing Spend, State of operation, and the historical profit generated by 50 startups. For a particular period, the profit earned by a company depends on several factors like how much time and money a company spends on research and development marketing and many more. So for a specific period predicting the profit of a company we need to convey a machine learning model with a dataset that contains old data about the profit accomplished by the company. The task of predicting profit is a main task for every business to set an achievable goal. For example, on marketing the business spends \$500 on marketing, it can't assume a profit of \$20,000. besides, the profit of a business depends on many other factors. Therefore, a company should set a goal that can be achieved. In the given, I'll lead you through the Python exercise of machine learning profit prediction. The major goal of this subsection is to make clear how key concepts are reused throughout the text. Scale-ups and start-ups are occasionally mentioned interchangeably. Both start-ups and scale-ups are the focus of this research. To avoid excessive repetition, these may alternatively be referred to as organisations, groups, realities, and topics of study. Independent variables are mostly explored as predictors in this study. These can be thought of as component or plain independent variables to avoid the word "continuous." The dependent variables are mostly discussed as criterion in this study. These can also be referred to as easily dependent variables, success criteria, or estimates of success to avoid repetition. Success: In this study, success is characterised as a binary variable. Based on the employed dependent variable, a start-up is categorised as either successful or unsuccessful.

➤ *Problem Statement:*

For both the administration and the stockholders, predicting the performance of start-up businesses is essential. It is difficult because there aren't enough relevant facts or reliable generic approaches. [Fig:1] Algorithms that use

machine learning can predict the future. The future success of start-ups must be predicted for both venture capital (VC) investors and startup businesses. Start-up organisations can change their development strategy and successfully snare openings by forecasting the future development of themselves and their competitors.

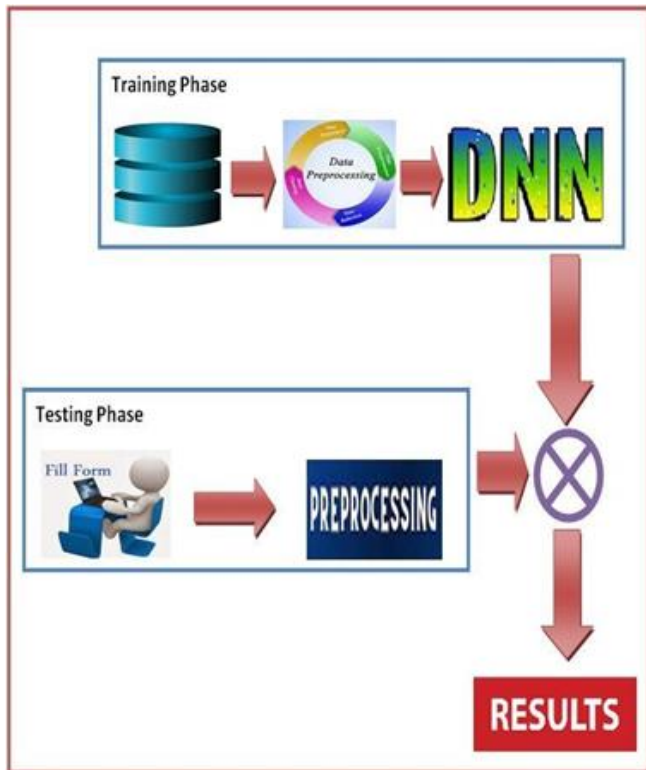


Fig 1 Startup Companies Profit Prediction

➤ *Technologies Used:*

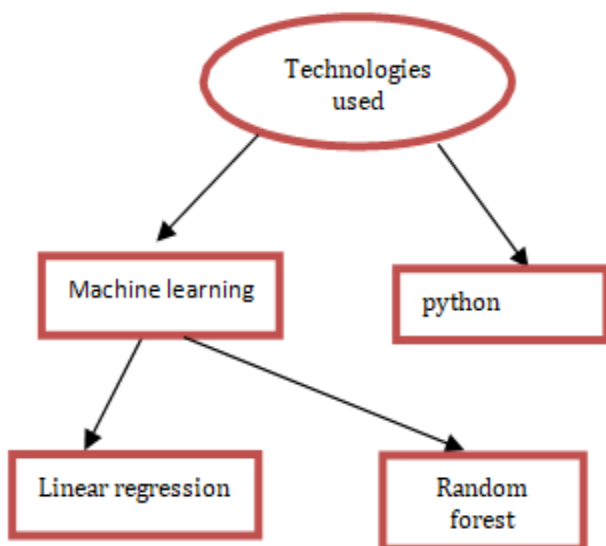


Fig 2 Technologies Used

• *Machine Learning:*

Machine learning (ML) is a field of inquiry devoted to comprehending and creating "learning" methods, i.e., methods that use data to enhance performance on a particular set of tasks. It is a part of artificial intelligence. A model

based on sample data, machine learning algorithms can be construct known as training data, in order to build predictions or resolve without being absolutely programmed to do so. In a wide variety of applications machine learning algorithms are used, such as in medicine, email filtering, voice recognition, cultivation, and computer vision, where it is hard or impractical to develop ordinary algorithms to perform the require tasks. It is subset it can closely related to computational statistics, which can focus on making predictions using computers, but not all machine learning is numerical learning. The study of mathematical reduction delivers methods, theory and application region to therange of machine learning. Data mining is a related plot of study, through unsupervised learning it can focus on exploratory analysis. Some fulfillment of machine learning use dataand neural networks in a way that acts the working of a biological brain. Across business problems its application, machine learning is also indicate to predictive analytics. Learning algorithms work on the base that strategies, algorithms, and consequences that worked well in the history are likely to continue working well in the future[Fig:2]. These consequences can be egregious, similar as "since the sun rose every morning for the last 10,000 days, it will presumably rise hereafter morning as well". They can be nuanced, similar as "X% of families have geographically separate species with color variants, so there is a Y% chance that undiscovered black swans live".

• *Linear Regression Model:*

In statistics, linear regression is a single approach for modelling the relationship between a scalar response and one or more analytical variables (also known as dependent and independentvariables).[Fig:3] The case of one analytical variable is called simple linear regression; formore than one, the process is called multiple linear regression. This term is specified from multivariate linear regression, where multiple activated dependent variables are predicted,rather than a one scalar variable.

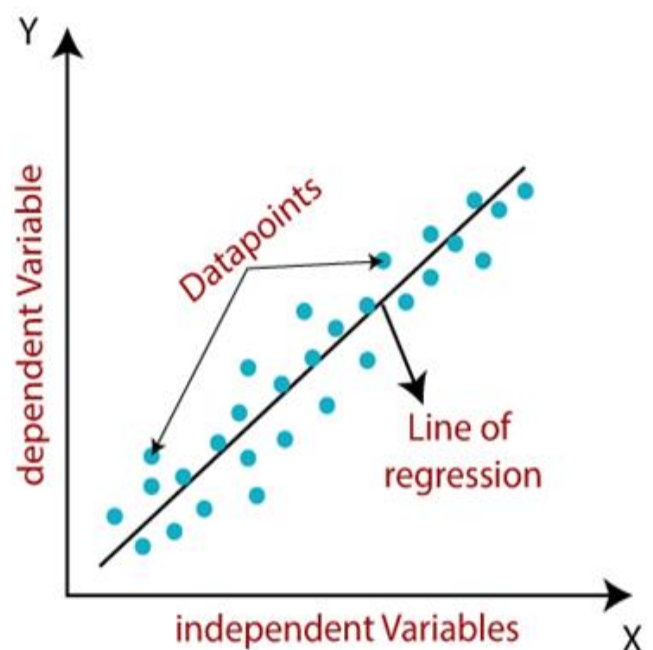


Fig 3 Linear Regression Model in MachineLearning

• **Random Forest:**

A random forest is created by growing many trees. However, individual trees are grown using only a subset of the available training data. For each tree, the method chooses a bootstrap sample of the training data (i.e. randomly chosen rows from the dataset). The method then chooses a random sample of the columns (features). The bootstrap sample with a subset of the available features is used to create an individual tree. [Fig:4] This process is repeated many times to grow a “random” forest of trees in which each tree is different. When it is time to make predictions, each tree in the forest is fed the testing data, and each tree makes a prediction. The majority vote of the forecasts made by the individual trees makes up the prediction of a random forest.

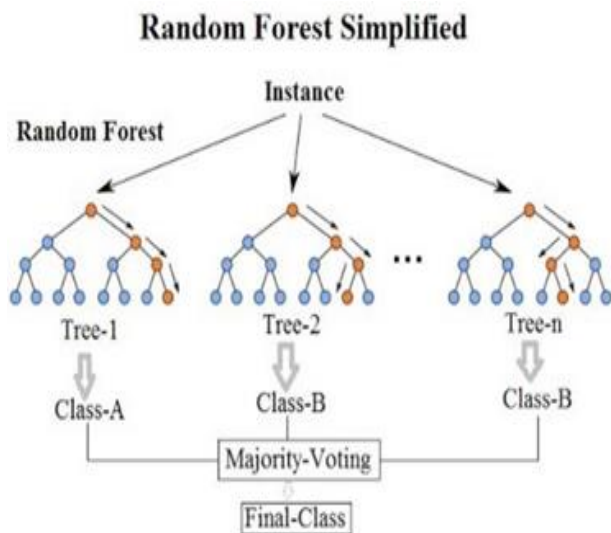


Fig 4 Random Forest in Machine Learning

• **Python:**

It is a well-known high-level programming language that is employed for general programming. The programming language was created by Guido van Rossum in the late 1980s, and it has since become one of the most popular in use worldwide. For helping developers create a wide range of apps, there are several libraries and frameworks available. Because of its well-known straightforwardness and ease of use, it is a good choice for both inexperienced and seasoned programmers. To expand the power of the language, Python provides a large range of modules that can be employed. There are several common modules:

✓ **NumPy:**

NumPy is a general-purpose library for managing arrays. Additionally, it provides tools for interacting with these arrays as well as a high-performance, greater dimensional array object. In scientific aggregation, Numpy is the primary tool for Python.

✓ **Matplotlib:**

There are a number of toolkits that are available to enhance the capability of Python's built-in matplotlib, which is not included in the Standard Libraries that are installed by default. Some of them can be downloaded separately, while others can be included in the matplotlib source code package but require additional resources.

✓ **Pandas:**

Join Pandas is a Python data analysis library. Wes McKinney founded pandas in 2008 in response to a demand for a robust and adaptable tool for quantitative research. Since then, pandas has become one of the most well-known Python libraries. It has a very vibrant contributor community.

✓ **Seaborn:**

It is a matplotlib-based Python data visualization package. It offers an advanced drawing interface for beautiful and practical statistics graphics.

➤ **Software Requirement Specification:**

• **Functional Requirements**

- ✓ **Operating system:** Modern operating system (windows 10)
- ✓ **Coding Language:** ML using python
- ✓ **Front-End:** Visual Studio 2012 Professional.
- ✓ **Data Base:** SQL Server 2008.

SRS is a captures complete description about how the system is expected to perform. At the conclusion of the requirements engineering phase, it is often approved. It defines how software system will interact with all internal modules, hardware, communication with each other programs and human user interactions with a wide range of real like scenarios.

➤ **Existing System:**

By using a single independent variable similar as the investment cost of a company’s design, the value of the dependent variable i.e., the profit of the company by the means of that design is roughly prognosticated. Linear retrogression makes use of a single independent variable to prognosticate the value of a dependent variable by developing a retrogression line along the given data and thereby prognosticating dependent variable using that retrogression line. There are some other ways viz the Bracket tree and RandomForest that makes use of a lot of dependent variable to prognosticate the value of the dependent variable and these ways works best for some of the given values but not for all. The main intention is to prognosticate the value of the dependent variable i.e., the value of the profit of the company grounded on the data of the company over the former times. So, from all the ways used before for the vaticination of profit an average from all those prognosticated values of the dependent variable is reckoned and made as the prognosticated dependent variable.

• **Drawbacks of Existing System:**

- ✓ **Unrealistic profit Targets.** A cast grounded on unrealistic hypothetical about the business terrain generally produces unobtainable profit targets.
- ✓ **Uses Too important Time**
- ✓ **Requires Financial Modelling chops**
- ✓ **Results unconnected to Strategies**
- ✓ **Creates Unrealistic prospects**

- ✓ *Linear regression makes use of only one independent variable and so results are less accurate.*
- ✓ *Data aren't fully consumed by a direct regression model.*

➤ *Proposed System:*

The main intention is to predict the value of the dependent variable i.e., the value of the profit of the company based on the data of the company over the previous years. So, from all the techniques used before for the prediction of profit an average from all those predicted values of the dependent variable is computed and made as the predicted dependent variable.

Training and testing are the two parts of the suggested system, which when combined result in a useful tool for any immovable object. The training phase includes the use of a database, data preprocessing, and a machine learning algorithm classifier. Complete the form to get it ready for testing. Both the training and testing phases are essential to

➤ *System Architecture:*

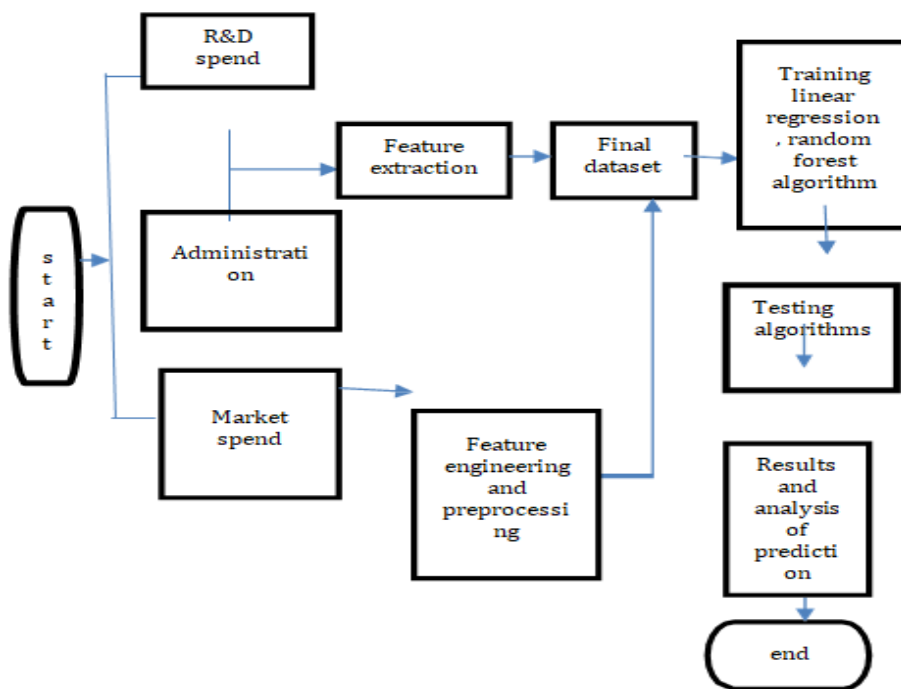


Fig 5 System Architecture

➤ *Future Scope:*

The best way to estimate the focus on the certain amount of money you are bringing home in a defined period, not agreement. funds to be honored later on. Ultimately, profit prediction is an in-depth analysis of past performance to help understand how much your business might bring in during the upcoming year.

II. CONCLUSION

This is how machine learning algorithms can be used to forecast a company's earnings for a specific time frame. These steps might assist a business in establishing an achievable goal. According to the assumption checking, the

obtaining an outcome. Everyone is aware that the business sector will offer us options to safeguard our future that will change our lives. Our suggested strategy is designed with profit estimation in mind. Because many young stars will prefer to start their own businesses, we developed our system for predicting start-up in this generation. Therefore, we use a database as an input. The database data includes information about marketing, administration, research and development, and the current state of the company. These prerequisites allow us to generate profit. Users' information can be gathered, and our system can then utilise that data to provide output that is useful to businesses and new startups alike.

• *Advantages of Proposed System:*

- ✓ It uses all the information provided to it to forecast the value of the independent variable.
- ✓ Hypothetically it is better than all the other existing machine learning algorithms.

model passes the homoscedasticity and no-multicollinearity tests but fails the residual normality test. The model with the lowest RMSE is the forward and backward model, which has an adjusted r2 of 0.94. This suggests that the predictor used in the model can interpret 94% of the data, while the remaining variables can interpret 6% of the data. We investigate whether a machine learning (ML) technique called classification trees can produce out-of-sample profitability projections that are more accurate than random walk forecasts. We are driven to use the ML technique.

REFERENCES

- [1]. Gao, Q., Xu, H., & Li, A. (2022). The analysis of commodity demand predication in supply chain network based on particle swarm optimization algorithm. *Journal of Computational and Applied Mathematics*, 400, 113760
- [2]. Greg Ross, Sanjiv Das, Daniel Sciro, Hussain Raza, *Capital VX: A machine learning model for startup selection and exit prediction*, *The Journal of Finance and Data Science*, Volume 7,2021.
- [3]. Daniel Cockayne, what is a startup firm? A methodological and epistemological investigation into research objects in economic geography, *Geoforum*, Volume 107, 2019.
- [4]. McNamara, Sue; McNamara, Anne (2019). Authentic simulated startups: Bringing the real world into the classroom. *Journal of Education for Business*
- [5]. P. Hajek and M. Z. Abedin, "A Profit Function-Maximizing Inventory Backorder Prediction System Using Big Data Analytics," in *IEEE Access*, vol. 8, pp. 58982-58994, 2020, doi: 10.1109/ACCESS.2020.2983118.
- [6]. Florida, Richard; King, Karen M. (2018). Urban Start-up Districts: Mapping Venture Capital and Start-up Activity Across ZIP Codes. *Economic Development Quarterly*, (), 089124241876373
- [7]. Jung, Sang Hoon; Jeong, Yong Jin (2020). Twitter data analytical methodology development for prediction of start-up firmsâ social media marketing level. *Technology in Society*.
- [8]. Arroyo, Javier; Corea, Francesco; Jimenez-Diaz, Guillermo; Recio-Garcia, Juan A. (2019). Assessment of Machine Learning Performance for Decision Support in Venture Capital Investments. *IEEE Access*
- [9]. Deller Mann, Dominik; Lipusch, Nikolaus; Ebel, Philipp; Leimeister, Jan Marco (2018). Design principles for a hybrid intelligence decision support system for business model validation. *Electronic Markets*
- [10]. Ramanjaneya Reddy M." EFFECTIVENESS OF SOCIAL MEDIA CONNECTED WITH CONTENT FOR START UP'S" 2022.
- [11]. Chih-Fong Tsai., Yu-Chieh Hsiao, "Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches", *Decision support systems*, 50,2010
- [12]. Filder.R. Hastings, "The organization and Improvement of the market Forecasting", *Journal of operation Research society*.45, 1994

BIOGRAPHIES

Dr. Sk. Mahaboob Basha is presently working as Professor in the department of Information Technolgy at NRI Institute of Technology, Vijayawada. He received his M.Tech degree from Jawaharlal Nehru Technological University, Kakinada (JNTUK) and Ph.D in Computer Science and Engineering from Acharya Nagarjuna University (ANU). He has published over 10 research papers in International Conferences and Journals. He has more than 20 years of experience in teaching.



Ch.sirisha is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. she donea summer internship on profit prediction.



R. manikanta Kumar is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. he done a summer internship on profit prediction.



P. Ramya Sri is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. she donea summer internship on profit prediction.