International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

Business Analysis using Machine Learning

Daya Mary Mathew[1], Lokesh J[2], Parthasarathy V[3]

[1]*Assistant Professor, Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai, Tamilnadu, India.*
[2-3]*Undergraduate student, Computer Science and Engineering, Rajalakshmi Engineering College, Chennai, Tamilnadu, India .*

Abstract - *The description and overview of the business plan for any existing retail shop or franchise is an elaborate description of the entire business. It describes what exactly the business does, how it does it (sells products in a shop-location of shop, area of shop, locality of shop, etc how it meets consumer wants, the specialized market or specific clients it serves, how it compares to competitors, what measures it takes, the corporate structure it follows, and so on. For demand forecasting and pricing, external data such as rival prices and weather conditions might be employed. Market and competition analyses are both included in the analysis. Market research and competitive analysis, as previously said, are critical to the success of a firm and must be conducted and documented in the business plan. The vital aspect to consider is choose the products that the shop should sell. The products to be sold at the venue must be decided in detail and mentioned in the business plan. Through machine learning will we analyze how much the quantity sold and what product sold at what time therefore this will be helpful to identify the profit in business.*

Key Words: competitive analysis, machine learning, analysis, business

1.INTRODUCTION

Business Analysis is the set of tasks, knowledge, and strategies required to become aware of commercial enterprise desires and decide answers to commercial enterprise problems. Although, the overall definition is similar, the practices and approaches can also range in numerous industries. In the Information generation industry, answers regularly consist of a structures improvement component, however, may include technique development or organizational change. Business evaluation can also be carried out to recognize the situation of a business enterprise or to function a foundation for the identity of commercial enterprise desires. In maximum cases, however, commercial enterprise evaluation is carried out to outline and validate answers that meets commercial enterprise desires, goals, or objectives.

2. RELATED WORK

1. Using a literature study and bibliometric analysis, proposed the research findings for an examination of the presence and evolution of the term Business Process Management (BPM) from 2000 to 2020. The goal of this study was to assess the number and quality of empirical support supporting the usage of this technology in businesses. This allowed the researchers to recognize and identify this discipline as an important study subject with a lot of potential for helping organizations achieve strategic alignment between business and ICT in the future.

2. The New Retail Business Analysis and Modeling was proposed. One of the most effective modes has been innovative retail. It is powered by information technology (big data, Internet of Things, artificial intelligence, etc.) and focuses on the user experience, as opposed to traditional business modes. It also reconstructs the essential features of online and offline commerce to create a new business mode. As a result, current business modelling tools are ineffective for analyzing and describing the new retail mode.

3. Contextual data has become a valuable asset in maximizing the value of information systems. The location of an event is an important type of context information that refers to where it takes place. The use of location-based analytics tools to enhance decision-making processes in company contexts is critical for business growth. However, following a thorough examination of the literature, we discovered that researchers and practitioners still lack a comprehensive characterization of location-based data analytics systems that have been successfully applied to business operations. This study summarizes the findings of a systematic literature review (SLR) in which we evaluated 168 location-based and business-oriented analytics solutions published between 2014 and 2019.

4. Proposed an illustration of the most recent results for a single business function's criticality ranking decision support classifier. The validated classifier is part of the business continuity points approach, which is based on the use case points method for measuring software complexity and assesses the recovery difficulty of an individual

business function. To determine precise recovery complexity factors of a given business function, the business continuity points technique is used. The business function criticality ranking, which is based on the recovery complexity factors, is one aspect of the strategy. In this research, we compare the outcomes of the rapid and comprehensive criticality ranking of a business function to determine the accuracy of the criticality ranking classifier.

5. In the face of various threat scenarios, measuring the security of corporate processes has become critical. A lot of research has been done on metrics for network security, software system security, attack severity, scenario evaluation, and so on over the previous two decades. Business impact analysis models and security maturity models, as well as well-established risk analysis approaches, are available at the process level. With the widespread use of IT to conduct business processes, it's become critical for chief information security officers to develop metrics for business process security in the context of the applicable threat scenario. Security Concern is a new security metric introduced in this research to analyze the security of business processes.

3. DATASET

The dataset used here is a collection of transaction data of a confectionery business which contains the following data:

- Date of transaction
- Time of transaction
- No of transactions
- Name of the product sold

Total no of transactions:21294

4. MACHINE LEARNING

Machine learning is a trendy issue for a variety of reasons, including the ability to extract deep insights, discover unseen patterns, and build high-performing prediction models from data without the need for explicit programming instructions.

If you're ever involved in a decision-making process involving the use of machine learning, how it can help you reach business and project goals, which machine learning techniques to apply, potential risks, and how to interpret the findings, you'll need this high-level expertise.

4.1. WORKING

These are the most typical machine learning challenges that one could encounter when attempting to tackle a machine learning problem. A list of machine learning algorithms that could be employed to solve these tasks is also mentioned under each task. If you think I've forgotten to mention something crucial, please leave a remark or make a suggestion.

The following are the important machine learning tasks that will be discussed in more detail later in this article:

Selection of features

- Regression
- Classification
- Clustering
- Querying with several variables
- Estimation of density
- Reduction of dimensions
- Matching and testing

5. MODELS

5.1 Random Forest

A random forest is made up of a large number of individual decision trees that work together as an ensemble, as the name suggests. The random forest creates a class prediction for each tree, and the class with the highest votes becomes our model's forecast.

The wisdom of crowds is the basic principle behind random forest, and it's a simple yet effective one.

The key is the low correlation between models. Uncorrelated models can provide ensemble forecasts that are more accurate than any of the individual predictions, similar to how low-correlation investments (such stocks and bonds) join together to build a portfolio that is larger than the sum of its parts.

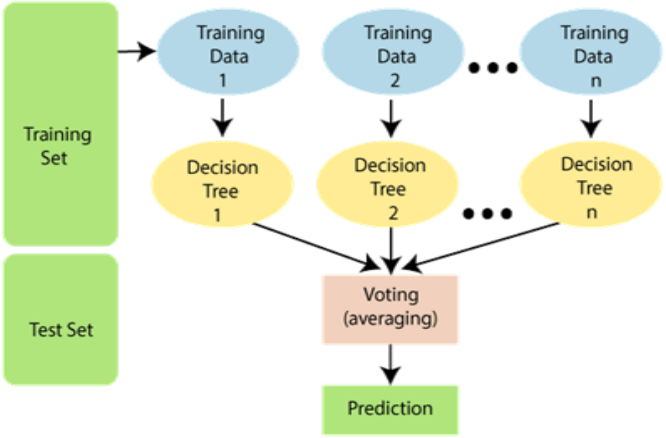

Fig 5.1.1 random forest

```
from sklearn.ensemble import RandomForestRegressor
regr = RandomForestRegressor(max_depth=2, random_state=0)
regr.fit(x_train, y_train)
rf_confidence = regr.score(x_test, y_test)
print("rf accuracy: ", rf_confidence)
```

```
rf accuracy:  0.8815693398026321
```

Fig 5.1.2 random forest accuracy

5.2 linear regression

Linear regression is one of the most straightforward and widely used Machine Learning techniques. It's a statistical strategy for predicting outcomes. Sales, salary, age, product price, and other continuous/real or numeric variables are all predicted using linear regression.

The linear regression algorithm demonstrates a linear relationship between a dependent (y) and one or more independent (y) variables, thus the name. Because linear regression depicts a linear relationship, it determines how the value of the dependent variable changes as the value of the independent variable changes.

The linear regression model provides a sloped straight line representing the relationship between the variables. Consider the below image:

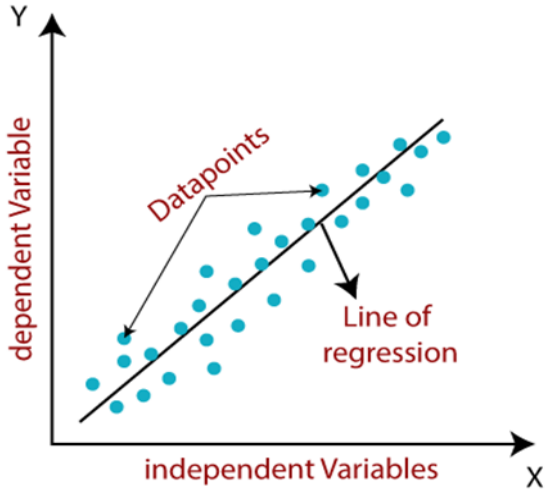

Fig 5.2.1 Linear Regression

Mathematically, we can represent a linear regression as:

Hello Java Program for Beginners

y= a0+a1x+ ε

Here,

Y=Dependent Variable (Target Variable)

X= Independent Variable (predictor Variable)

a0= intercept of the line (Gives an additional degree of freedom)

a1 = Linear regression coefficient (scale factor to each input value).

ε = random error

The values for x and y variables are training datasets for Linear Regression model representation.

```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(x_train, y_train)
lr_confidence = lr.score(x_test, y_test)
print("lr accuracy: ", lr_confidence)
```

```
lr accuracy:  0.8100714106137765
```

Fig 5.2.2 Linear Regression accuracy

5.3 Decision Tree

We may thoroughly analyze the probable repercussions of a decision using a decision tree.

It gives us a framework for calculating the worth of outcomes and the likelihood of achieving them.

It assists us in making the best judgments possible based on available data and best guesses.

In other terms, a decision tree is a hierarchical tree structure that may be used to divide a large collection of records into smaller sets of the same class using a set of basic decision rules. A decision tree model is a collection of rules for breaking down a large, heterogeneous population into smaller, more homogeneous, or mutually exclusive groups.

The classes' properties can range from nominal, ordinal, binary, and quantitative values; nevertheless, the classes must be of a qualitative nature, such as categorical, ordinal, or binary. In a nutshell, a decision tree generates a set of rules that may be used to identify a class based on the given data of characteristics and its class. A hierarchy of segments within a segment is created by applying one rule after another. The tree represents the hierarchy, and each node represents a segment. The members of the successive sets become increasingly similar as the divisions advance. As a result, recursive partitioning is the name given to the algorithm used to generate a decision tree. The algorithm is known as CART (Classification and Regression Trees).

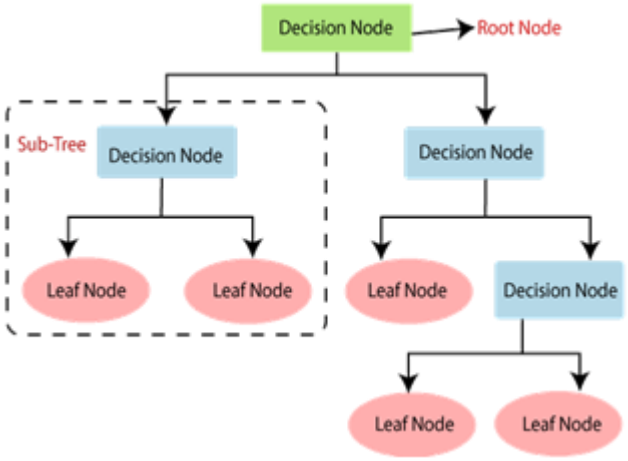

Fig 5.3.1 Decision Tree

```
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor()
regressor.fit(x_train, y_train)
dt_confidence = regressor.score(x_test, y_test)
print("dt accuracy: ", dt_confidence)
```

```
dt accuracy:  0.909262668692403
```

Fig 5.3.2 Decision Tree Accuracy

6. CONCLUSIONS

In present days many of the researchers were focusing on the concept of ML and AI for the things such as Business forecasting, analysis, approximation etc. our paper completely focus on the data regarding business mechanisms on the sales information using the concept of ML. AI is the main mechanism in increasing in terms of price as well as loyalty. For such procedures there are several features such as time series, loyalty, etc., to overcome such types of problems the things like time series, analysis, capturing historic data so-on came into the existence. For developing these concepts there are few forecasting algorithms such as linear regression, random forest, etc., were used. In our work we have taken the sales data regarding the pastry shop and their sales information. The overall performance of each product is calculated by the help of machine learning algorithms like linear regression, random forest, decision tree. With the produced results we are able to analyze the performance of each and every product sold by the pastry at different cycles of the day therefore enabling us to optimize the products according to their demand and supply.

REFERENCES

1.Uma Kumar, Kayvan M. Lavassani, Vinod Kumar, Bahar Movahedi.(2008). Measurement of Business Process Orientation in Transitional Organizations: An Empirical Study.

2.Yuyu Yin; Rong Zhang; Honghao Gao; Meng Xi.(2018).New Retail Business Analysis and Modeling: A Taobao Case Study.

3. Luis E. Ferro-Díez; Norha M. Villegas; Javier Díaz-cely.(2020). Location Data Analytics in the Business Value Chain

4.Athanasios Podaras.(2018). Measuring the Accuracy Levels Regarding the Dual Business Function Criticality Classifier

5.Preetam Mukherjee; Chandan Mazumdar.(2018). Security Concern" as a Metric for Enterprise Business Processes