International Conference on Innovative Research in Engineering, Management and Sciences
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 4 | Issue 9 | ISSN : 2456-3307

Business Practices using Machine Learning

Mounica.B [*1]

[*1] Information Science & Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India
premkumarmounica@gmail.com [1]

ABSTRACT

Busy life schedule leading to the growth of E-commerce companies. Many Challenge in development of business through web pages, apps and retail stores. Proper utilization of invested funds by analyzing costumer preferences and purchases from the previous marketing data. Business practices are very arduous to transmute especially in astronomically immense and mature companies. Introducing incipient ways of working is astronomically arduous, but making people work differently is virtually infeasible. There is immensely colossal cultural inertia in these companies and business industry in general. Utilization of Machine Learning is one of those transmutations that will make people work differently and will make business environments different in future.

Keywords : Machine Learning algorithms; Supervised Learning; Unsupervised Learning; Reinforcement Learning.

I. INTRODUCTION

By using the seaborn library the statistical graphs will be drawn between data sets(yearly amount spent ,time spent on app, time on website ,retail shop length)Finding the cross co-relating factor that is linear to the x and y factors. Using the cross co-relating factor we test and train the model with our own set of values and find the average co-ordinates between them .Each co-ordinates specify the values such has time on app, time on website and membership length. By which we decide to develop the business.

Multinational company that has both E-commerce and retail store which decides whether to focus their efforts on their mobile app experience or their website or to develop their retail stores.

The basic consumer data such has name, email, age, length of membership, yearly amount spent, time spent on app or website and internet used.

All this data will be collected from company has a sample data Processing the data involves the usage of libraries such has pandas, NumPy and seaborn. Pandas is a open source library providing high-performance, easy-to-use data structures and data analysis tools for programming languages. NumPy is a library for the adding support of large multi-dimensional arrays and matrices. along with mathematical functions to operate on arrays .Seaborn tool is a data visualization library works on matplot lib. It provides joint plot options to compare the linearity and co-relation factor.

A. Role of big data in business development

The true role played by big data in business development is important and far-reaching, especially when it comes to long term client satisfaction. In the best of all possible worlds, companies will use the data they collect to improve their products and the customer experience .As we invite more connected things into our lives — from smart thermostats to Apple watches and fitness trackers. These are just the top four impacts I soothsay immensely colossal data will have on businesses of all types in the near future.

Even the smallest businesses generate data these days. If the company has a website or a social media related or accepts credit cards etc., even a one-person shop has data it can collect on its customers, its user experience, web traffic, and more. This means companies of all sizes need a strategy for big data and plan how to collect, use, and protect it.

Business development is a process of developing a long term business model. It governs the use of analytics and focuses on every stage of the process, from developing a strategy to executing it.

The goal is to create a successful business model for all stakeholders. Of course, shareholders need to know that the model will provide great compensation for them. However, business development must also focus on the needs of customers, employees and others affected by the business.

Big data is a field of engineering that deals with large datasets which follows the ETL method of capturing the data from one data environment, extracts it, transforms and loads it into another data environment in an effective manner. Big data is an essential technology when efficiency and accuracy is of highest concern. It integrates and analyses massive amounts of data to reduce the cost of failure and improve the system. It is also a database for correlating the collected data and the user can access a part of the stored information.

B. Machine Learning algorithm

There are many Machine Learning Algorithms helps for analysis. The method of how and when you should be using them. By learning about the List of Machine Learning Algorithm you learn furthermore about AI and designing Machine Learning System.

Machine Learning(ML) can be explained as automating and improving the learning process of computers based on their experiences without being actually programmed i.e. without any human assistance. The process starts with feeding good quality data and then training our machines(computers) by building machine learning models using the data and different algorithms. The cull of algorithms depends on what type of data do we have and what kind of task we are endeavoring to automate.

These are the most paramount Algorithms in Machine Learning. If you are vigilant of these Algorithms then you can utilize them well to apply in virtually any Data Quandary. Data analyst and the Machine Learning developers utilize these Algorithms for engendering sundry Functional Machine Learning Projects. Then comes the 3 types of Machine Learning Technique or Category which are utilized in these Machine Learning Algorithms. The three categories of these Machine Learning algorithms are:

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

1).The supervised Learning method is used by maximum Machine Learning Users. It is called Supervised Learning because the way an Algorithm's Learning Process is done, it is a training data set. And while using Training dataset, the process can be thought of as a teacher supervised the Learning Process. The correct answer is known and stored in the system already. The algorithm Predict Data that is in Training Process and gets the results[1][2][3].

Supervised learning is the learning of the model where with input variable (say, x) and an output variable (say, Y) and an algorithm to map the input to the output

That is, Y = f(X)

Why supervised learning?

It is called supervised learning because in this process (from the training dataset) can be thought of as a mentor who is supervising the entire learning process. Thus, the "learning algorithm" iteratively makes predictions on the training data and is corrected by the "mentor", and the learning stops when the algorithm achieves an acceptable level of performance(or the desired accuracy).

Example of supervised learning:

Suppose there is a basket which is filled with some fresh fruits, the task is to arrange the same type of fruits at one place.

Also, suppose that the vegetables are onion, tomato, chilli, ginger. Suppose we already knows from their previous work (or experience) that, the shape of each and every vegetable present in the box so, it is easy for people to arrange the same type of vegetable in one box.

Here, the previous work is called as training data in Data Mining terminology. So, it learns the things from the training data. This is because it has a response variable which says y that if some vegetable has so and so features then it is onion, and similarly for each and every vegetable. This type of information is deciphered from the data that is used to train the model.

There are two types of supervised learning problems. These Supervised quandaries can be further grouped into regression and relegation quandaries.

- Classification Quandaries: Relegation quandary can be defined as the quandary that brings output variable which falls just in particular categories, such as the "red" or "blue" or it could be "disease" and "no disease".
- Regression: A regression quandary is when the output variable is a real value, such as "dollars" or it could be "weight".

Unsupervised Learning: Unsupervised learning is where only the input data (say, X) is present and no corresponding output variable is there.

Why Unsupervised Learning?

The main aim of Unsupervised learning is to model the distribution in the data in order to learn more about the data.It is called so, because there is no correct answer and there is no such mentor(unlike supervised learning). Algorithms are left to their own devises to discover and present the interesting structure in the data ,it works with unlabeled data set.

Unsupervised learning is that algorithm where you only have to insert/put the input data (X) and no corresponding output variables are to be put.

The major goal for the unsupervised learning is to avail model the underlying structure or maybe in the distribution of the data in order to avail the learners learn more about the data.

Unsupervised learning quandaries can even be grouped ahead into clustering and sodality quandaries[4][5][6].

1. Clustering: A clustering is that problem which indicates what you want to discover and this helps in the inherent groupings of the data, such as grouping the customers based on their purchasing behavior.
2. Association: An association rule is termed to be the learning problem. This is where you would be discovering the exact rules that will describe the immensely colossal portions of your data. Example: People who buy X are withal the one who inclines to buy Y.

C. Linear Regression

Before knowing what is linear regression, let us get ourselves accustomed to regression. Regression is a method of modeling a target value predicated on independent presages. This method is mostly utilized for forecasting and ascertaining cause and effect relationship between variables. Regression techniques mostly differ predicated on the number of independent variables and the type of relationship between the independent and dependent variables.

Simple linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and dependent(y) variable. The red line in the above graph is referred to as the best fit straight line. Predicated on the given data points, we endeavor to plot a line that models the points the best. The line can be modeled predicated on the linear equation .As the name indicates this already, linear regression is well known to be an approach for modeling the relationship that lies in between a dependent variable 'y' and another or more independent variables that are denoted as 'x' and expressed in a linear form. The word Linear indicates that the dependent variable is directly proportional to the independent variables. There are other things that are to be kept in mind. It has to be constant as if x is increased/decreased then Y also changes linearly. Mathematically the relationship is based and expressed in the simplest form as: This isy = Ax +B. Here A and B are considered to be the constant factors. The goal hidden behind the Supervised learning using linear regression is to find the exact value of the Constants 'A' and 'B' with the help of the data sets. Then these values, i.e. the value of the Constants will be helpful in predicting the values of 'y' in the future for any values of 'x'. If there is a single and independent variable it is called as simple linear regression and if there is more than one independent variable, then this process is called multiple linear regression.

The Mundane Least Squares Regression or call it mundane least squares (OLS). The linear least squares. When we consider the statistics, this is a method where we estimate the unknown parameters. This is kenned as the linear regression model, it comes with the goal which minimizes the differences of the observed replications in some arbitrary dataset.

Withal, minimizes the replications that are very well soothsaid by the linear approximation of the data (visually this can be optically discerned as the sum, which is of the vertical distances falling in between each data point in the set and the corresponding points on the regression line – it is observed that the more minute the differences are, the better would be the model that fits the data).The resulting estimator can be expressed in the form of a simple formula, especially when this falls in the case of a single regressor and is on the right-hand side. The OLS estimators are kenned to be authentically consistent whereas the regresses are exogenous and there lies no impeccable multicollinearity, and this remains optimal in the class of the linear equitable estimators. While there are errors, these are homoscedastic and serially uncorrelated. Under these conditions, there is a method of OLS. It provides with the minimum-variance, there is a mean-unbiased estimation, here the errors would have finite variances. Under these additional assumptions, there are errors that could be normally distributed. The OLS algorithm is the maximum likelihood estimator[7][8].

II. LINEAR REGRESSION USED FOR BUSINESS DEVELOPMENT

By using the seaborn library the statistical graphs will be drawn between datasets(yearly amount spent , time spent on app, time on website , retail shop membership length) .Finding the cross co-relating factor that is linear to the x and y factors. Using the cross co-relating factor we test and train the model with our own set of values and find the average co-ordinates between them. Each co-ordinates specify the values such as time on app, time on website and

membership length. By which we decide to develop the business. After that we get large points in x and y axis so, to get the exact average of the customer outcome we use linear regression Compared to any other model linear regression gives more accurate results for the customer detail outcome[9][10].

Sample code is given below

```
customers = pd.read.csv("Ecommerce Customers")
customers.head()
customers.describe()
sns.jointplot(x=&#39;Time on Website&#39;,y=&#39;Yearly Amount Spent&#39;,data=customers)
sns.jointplot(x=&#39;Time on App&#39;,y=&#39;Yearly Amount Spent&#39;,data=customers)
sns.jointplot(x=&#39;Time on App&#39;,y=&#39;Length of Membership&#39;,kind=&#39;hex&#39;,data=customers)
sns.pairplot(customers)
sns.lmplot(x=&#39;Length of Membership&#39;,y=&#39;Yearly Amount Spent&#39;,data=customers)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=101)
predictions = lm.predict( X_test)
print(&#39;MAE:&#39;, metrics.mean_absolute_error(y_test, predictions))
print(&#39;MSE:&#39;, metrics.mean_squared_error(y_test, predictions))
print(&#39;RMSE:&#39;, np.sqrt(metrics.mean_squared_error(y_test, predictions)))
sns.distplot((y_test-predictions),bins=50);
coeffecients = pd.DataFrame(lm.coef_,X.columns)
coeffecients.columns = [&#39;Coeffecient&#39;]
coeffecients
```

III. RESULTS

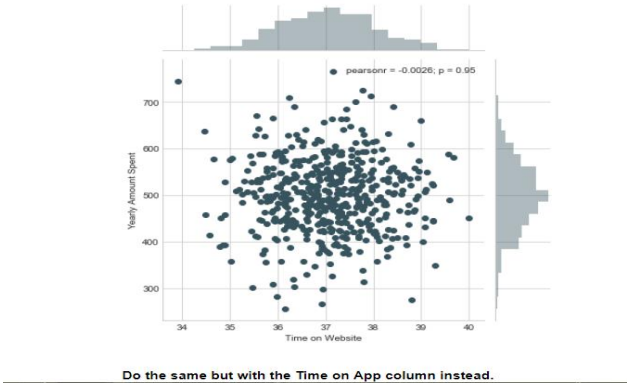

Figure 1. Time spent on Website

In figure.1 Use seaborn to create a joint plot to compare the Time on Website and Yearly Amount spent columns. Here we see the total duration spent by the customer on website and the amount spent per annum on the website.

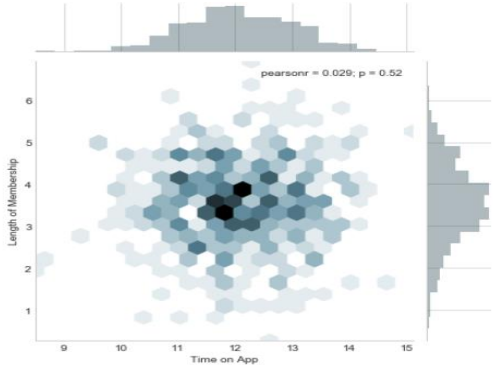

Figure 2. Time spent on App

In figure.2 we have a graph between Time on App and Length of membership. Here we see the total of customers spending time on app and their membership of how long they've been part of the company as a customer

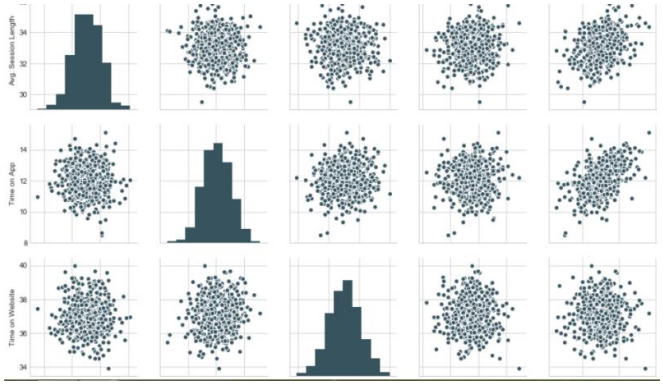

Figure 3. Relationships of the entire data set

In Figure.3 we come across the relationships of the entire data set. Here we are co-relating all the columns of the customer details to each other in the

graphs to find out the accurate co-relating out-put graph.

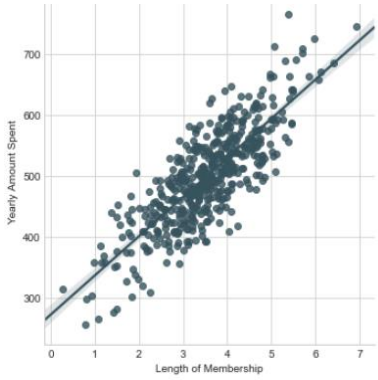

Figure 4. Yearly Amount Spent vs. Length of Membership

In the above Figure.4 the graph is between Yearly Amount Spent vs. Length of Membership. We get this graph in the final output because it's the most accurate graph out of all the co-relating graphs and we get the results based on this graph.

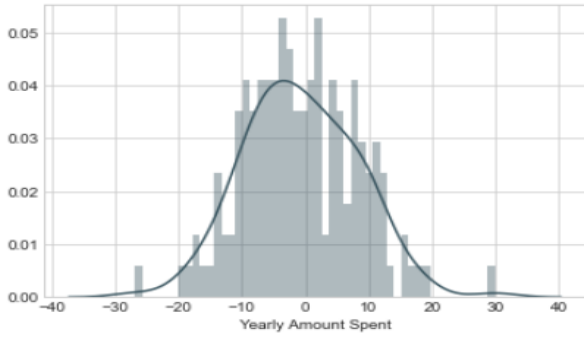

Figure 5. Time spent on App and website

	Coeffecient
Avg. Session Length	25.981550
Time on App	38.590159
Time on Website	0.190405
Length of Membership	61.279097

Figure 6. Results of the time on App and the time spent on the website

This is the final result. These results are based on final output graph and by this we decide on what can we invest in and according to the above results the time on App spent is greater than the Time spent on the website, by this accuracy we conclude that investment on App is better than investment on website.

IV. CONCLUSION

Multinational company that has both E-commerce and retail store which decides whether to focus their efforts on their mobile app experience or their website or to develop their retail stores .And to find the best way to advertise their products in most efficient way through these app and websites. Investment funding in E-commerce (app/web page) or retail store can be determined and is the best way of advertisement.It Increase the research efficiency in investment department. This project limits the loss of the investment. And it is helpful to detect the best way of advertising of the product. And have satisfied clients.

V. REFERENCES

[1] Corchado, E., Graña, M., & Wozniak, M. (2014).A survey of multiple classifier systems as hybrid systems Information Fusion, 16, 3-17.

[2] X. Y. Liu, 2012. C. H. Gao, P. Li, "A comparative analysis of support vector machines and extreme learning machines", Neural Netw., vol. 33, pp. 58-66.

[3] S. Chen, C. Cowan, P. Grant, 1991. "Orthogonal least squares learning algorithm for radial basis function networks", IEEE Trans. Neural Netw., vol. 2, no. 2, pp. 302-309

[4] R. Zhang, Y. Lan, G.-B. Huang, Z.-B. Xu, 2012. "Universal approximation of extreme learning machine with adaptive growth of hidden nodes", IEEE Trans. Neural Netw. Learn. Syst., vol. 23, no. 2, pp. 365-371.

[5] E. D. Karnin, 1990. "A simple procedure for pruning back-propagation trained neural networks", IEEE Trans. Neural Netw., vol. 1, no. 2, pp. 239-242

[6] C. Cortes, V. Vapnik, 1995. "Support vector networks", Mach. Learning., vol. 20, no. 3, pp. 273-297

[7] J. A. Suykens, J. Vandewalle, 1999. "Least squares support vector machine classifiers", Neural Process. vol. 9, no. 3, pp. 293-300.

[8] S. Tong, D. Koller, 2002. "Support vector machine active learning with applications to text classification", J. Mach. Learn. Res., vol. 2, pp. 45-66.

[9] L. Zhang, W. Zhou, L. Jiao, 2004. "Wavelet support vector machine", IEEE Trans. Syst. Man Cybern. B Cybern., vol. 34, no. 1, pp. 34-39.

[10] D. E. Rumelhart, G. E. Hinton, R. J. Williams, 1986. "Learning representations by back-propagating errors", Nature, vol. 323, no. 6088,pp. 533-536.