***BGMI WINNER RANKING PREDICTION USING RANDOM FOREST SCALABLE MACHINE LEARNING PLATOFORM***

*Harsh Chheda1*

*Department of Computer Science*

*S.K Somaiya College , Somaiya Vidyavihar University*

*Mumbai , India*

[*harshc3294@gmail.com*](mailto:harshc3294@gmail.com)

**Abstract-- The next generation is heavily influenced by gaming, and with this influence comes progress, whether it be in terms of aesthetics, music, gameplay, engagement, or ranking. The ranking fosters healthy rivalry by letting players know where they stand globally. This increases their level of competition and helps them rise above the competition. There are a ton of gamers online every day playing the multiplayer game BGMI, which is supported on a number of different platforms. Online video game BGMI has recently gained a lot of popularity among young people. One of the most crucial aspects of this game is final rank, which represents a player’s performance. This essay focuses on estimating the player’s ultimate ranks based on their aptitudes and capabilities. We have adjusted the number of features utilized in the model using the correlation heatmap.** **Random Forest is a machine learning algorithm that was employed in this study. The MAE (Mean Absolute Error) was used for the ranking of different players in the metric evaluation of all these algorithms.**

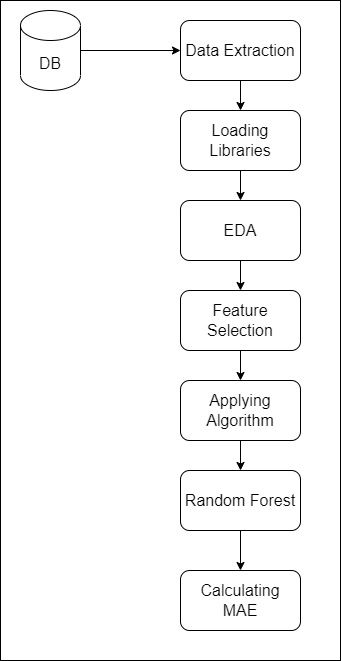
***Keywords: bgmi, Machine Learning, Random Forest Method.***

1. **INTRODUCTION**

BGMI, is a multiplayer game that can be played on a variety of operating systems Android, and iOS. Three different modes are available in the game: Classic, Arcade, and EvoGround. The user will be given a variety of maps in classic, including those for Erangel, Miramar, Sanhok, and Vikendi. War, Mini-Zone, Quick Match, and Sniper Training are available in arcade mode. The ranking difficulty increases with the amount of players, which totals 555 million across all platforms. In a Battle Royale contest, 100 players compete in a single match with only 1 winner (who can have Chicken Dinner). It is challenging to evaluate these athletes according to many characteristics because some players may receive more than one ranking that is similar. In order to anticipate the ranking of players based on the trained model, machine learning and deep learning might be useful in this situation by examining numerous attributes and comprehending the similarities between each. The dataset was obtained from BGMI, an open-source platform for gathering data pertaining to various use cases, where they combined various matches (SOLO, SQUAD, and DUO) with all different sets of attributes for understanding the use case. In this study, we employed K-Nearest Neighbors (KNN) [1], Gradient Boosting Regression (GBR), Decision Tree [2], Multiple Linear Regression (MLR), and other previously used algorithms to estimate the rank of the players or teams in BGMI using a dataset. The Random Forest algorithm, which has an MAE of 0.02 and the maximum accuracy of almost 95%, is what we find. We also demonstrate that the model's performance can be maintained using only 8 features (the top 8 features of the correlation heatmap). However, this decrease in the number of features speeds up the empirical runtime of this approach by about 1.5 times. We have also numerically demonstrated that the performance of ML algorithms is significantly impacted by further lowering the amount of features. The rest of the paper is organized as follows.

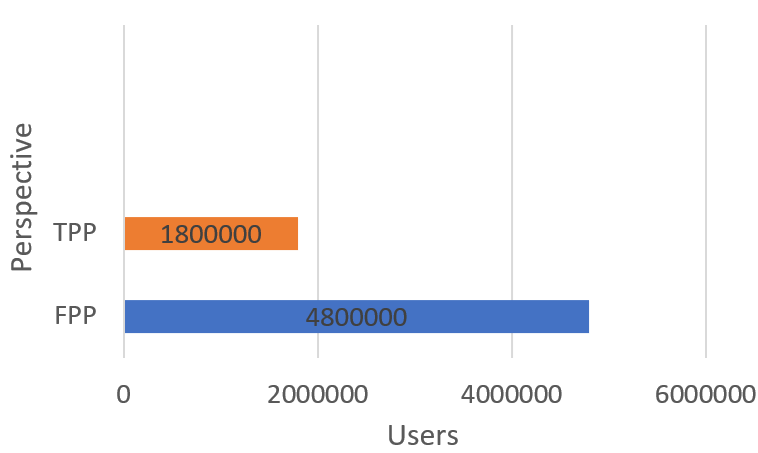
1. **LITERATURE REVIEW**

This research paper's experimental portion includes a number of tests and findings that were attained through iterations. The datasets of 29 attributes, which have a size of 446966 for the training dataset and 1934174 for the testing dataset, were used in this research paper's experimental outcomes. These datasets were subjected to multiple algorithm iterations and exploratory data analysis. The experiment's system architecture can be seen in the image below.



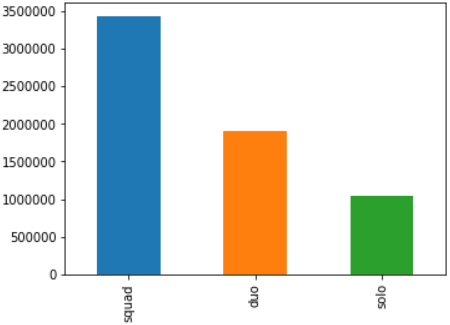
**Fig 1. Architecture for BGMI**

Exploratory data analysis is carried out in order to comprehend the visual representation of characteristics connected to the target variables and to improve judgments based on the significance of features. FPP stands for First Person Perspective, and TPP stands for Third Person Perspective. These perspectives change the user's field of view from wide to narrow based on focusing directly through weapons from a different perspective. The first EDA was performed to understand the number of players playing in each perspective.



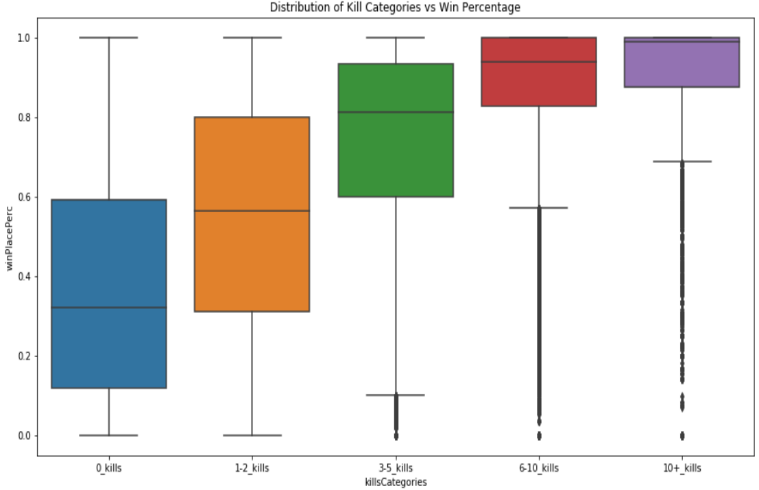
**Fig 2. TPP vs. FPP User Perspective**

As observed in Fig, the majority of players use the FPP perspective while only a small minority use the TPP due to the TPP's extremely constrained screen, which makes aiming challenging. The second EDA was conducted in order to comprehend the category in which players tend to play. PUBG offers three categories: SOLO, DUO, and SQUAD, which can be shown in the Fig. below.



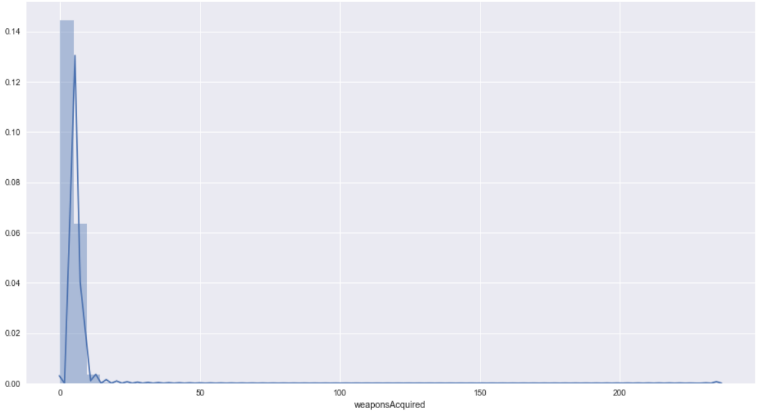
**Fig 3. Solo vs. Squad vs. Duo Users**

As seen in Fig, the majority of players fall into the Squad category, followed by Duo, and the smallest number of players play solo.



**Fig 4. Kill Categories vs. Win Place Prediction**

The distribution of kills in relation to win prediction is shown in the box plot in the aforementioned Fig. It can be observed that win prediction rises as kills rise; roughly 3-5 kills favor a 0.8 percentile of win prediction, and more than 10 kills in a match can result in 100% win rates. Another EDA was carried out to plot the histogram for weaponry that players in the game acquired (total of solo, duo & squad). This graph displays the player's total number of weapons along with their distribution. As can be seen from the following graph, each user had an average of 8 weapons, ranging from 0 to 15.



**Fig 5. Weapons Acquired vs. the Prediction Place**

With the aid of a correlation matrix, the final EDA was carried out to comprehend the correlation of attributes pertaining to each. The dataset's correlation matrix is displayed in Fig. below.

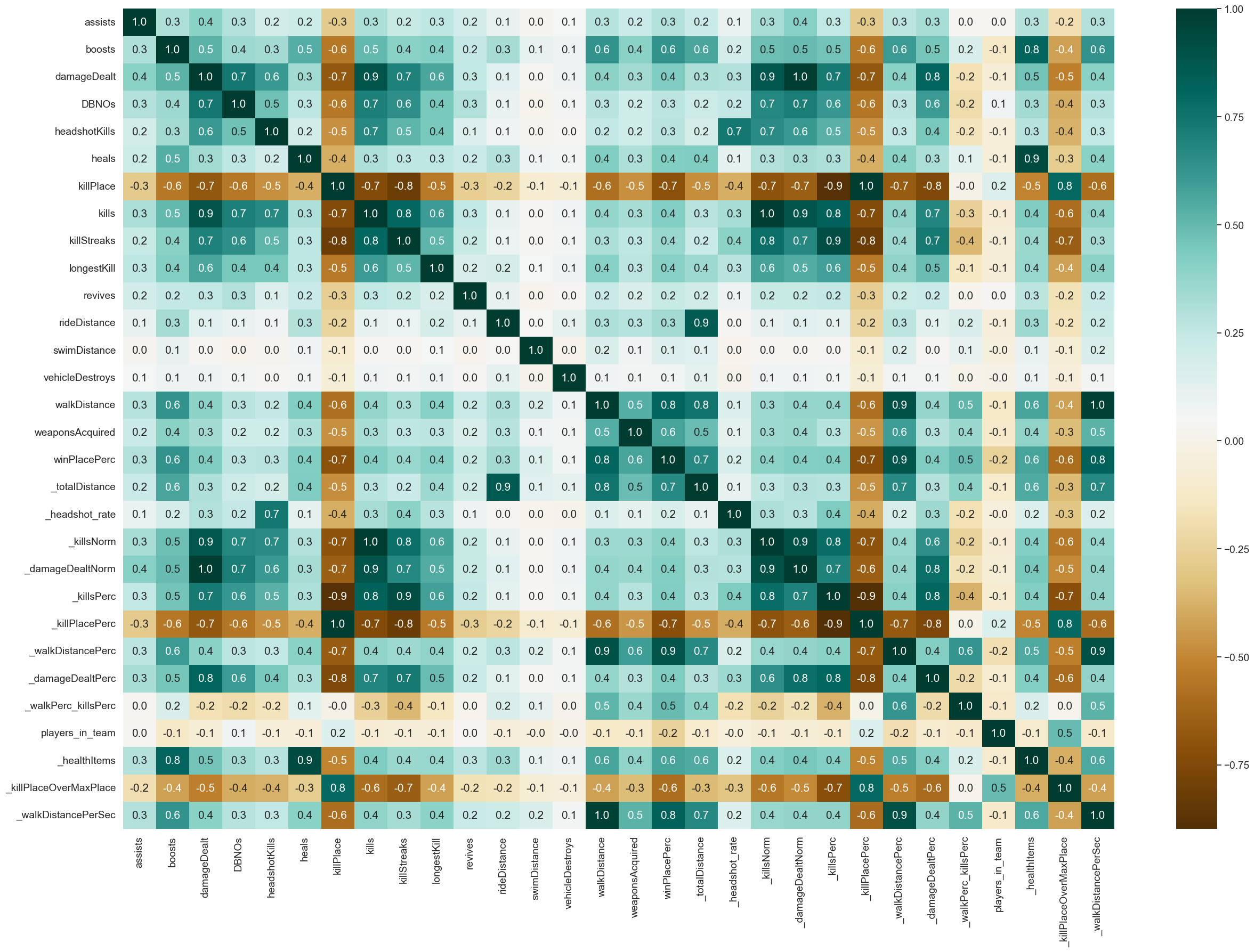
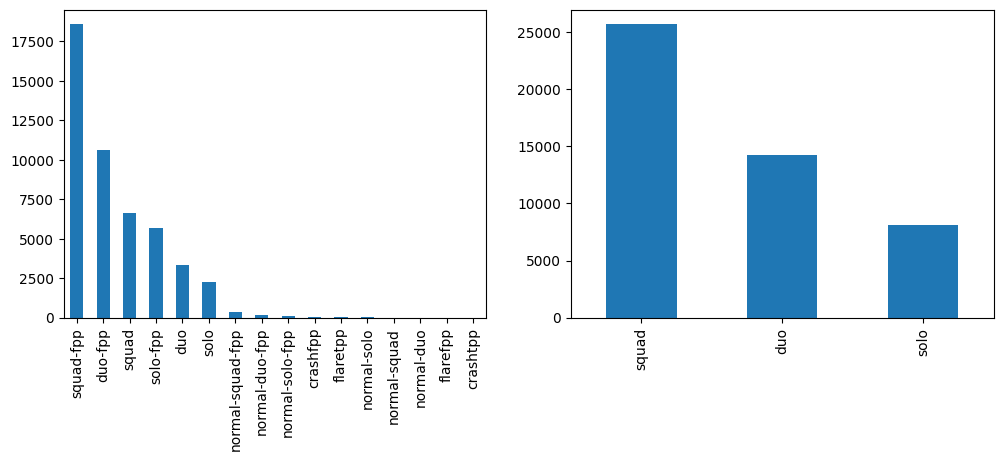
\

Fig 6. Correlation Matrix for BGMI

The correlation matrix displays the degree of correlation between attributes in datasets, indicating whether they are highly connected or not. Positive 1 (+1) in Fig. 6 above indicates a high correlation, whereas negative 1 (-1) indicates a low correlation. With the help of this EDA, significant data was extracted through visualization. The following phase was to identify the key features that would be used to train the model. If these features were chosen based on relevance, they would perform better than features that were randomly given to the model.

1. **METHODOLOGY**
2. **Data Cleaning**

For SOLO, DUO, and SQUAD match types, the maximum number of participants is 1, 2, and 4, respectively. However, the dataset has divided the match types into 16 distinct types, each of which is a variation of the three basic categories. The following is how we divided the number of players into the appropriate match type (Fig. 7):



**Fig 7. Mapping of different match types into core match types.**

Game players' type

j = ∑i players in the game type ji … (1)

where ji are the different sub formats of the primary game from j

Additional aberrant data have been eliminated from the dataset. One or more of the following conditions must be met for a data to be removed: I More players are in a team than are permitted for a given match type; (ii) Players that have no kills, haven't moved much, and haven't picked up any weapons are (perhaps offline).

1. **Feature Engineering and Selection**
2. **FUTURE SCOPE**

Some of the changes can be done into future are:

* Artificial Intelligence Mechanism can be added.
* Other the college subject , new courses can be added apart from college subjects
* Live lecture can be added
* Online practical section can be added

1. **CONCLUSION**

The goal of the project is to rate BGMI players based on machine learning and deep learning algorithms, as well as perform EDA for more thorough dataset analysis. The algorithms employed in this study, Random Forest MAE (Mean Absolute Error), were calculated to see which one specifically fit the large dataset the best. The Random Forest Algorithm worked well, with an MAE value of 0.029 for the training dataset and 0.025 for the testing dataset, respectively, with n estimators=50 and max features=0.5, according to this evaluation, which was done for both the training and testing datasets. The most kills ever recorded in a single game were 72 kills in one match, with the average player killing 0.94 players, and 99% of players having 7.0 kills or fewer. The Random Forest Algorithm, which had the lowest error value for testing data, was the best algorithm out of all of them.

**ACKNOWLEDGEMENT**

**I pay appreciation to my supervisor Mrs. Swati Maurya for guidance, support, patience and understanding and comprehension throughout the research time frame, likewise my genuine appreciation goes to S.K. Somaiya college for giving me this chance.**

**REFERENCES**

[1] Computational Statistics and Data Analysis, volume 38, Issue 4, Pages 367-378,Elsevier, 2002 .

[2] Annals of Translational medicine, Vol 4, No 11 , 2016