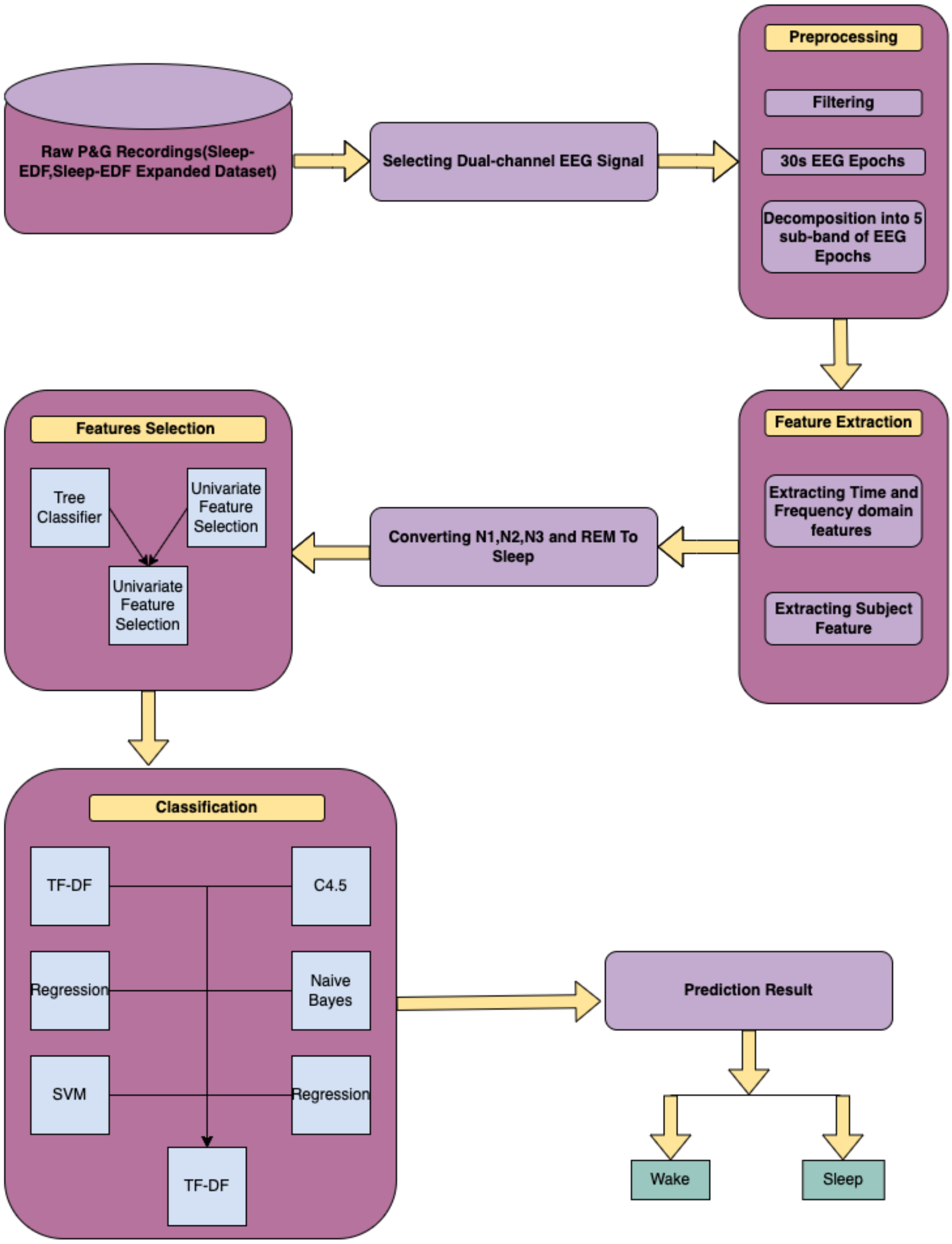

Figure 1: Structural framework of the proposed automated sleep staging system.

3. Methodology:

The flow diagram of the proposed method is given in Figure 1. The proposed method have been explained in the following subsections.

3.1.1 Data Acquisition:

The dataset used in this proposal is collected from a sleep center named Haaglanden Medisch Centrum. It was collected during 2018 and published very recently on 1st July, 2021. The dataset includes Whole-night Polysomnographic(PSG) sleep recording of 154 people (88 Male, 66 Female) with Mean Age of 53.8 ± 15.4. Each recording Contains a raw signal file (.edf) and a sleep scoring file (.txt) that gives a sleep score for an epoch of 30 sec.Each PSG recording contained several types of signals including EEG, EOG, EMG and ECG. We have used three channels F4,C4 and O2 EEG signals has 108452 rows of data.

3.1.2. Wavelet Package Decomposition

Five level one-dimensional wavelet decomposition of each epoch is done wavelet. The five level of decomposition produced six different stages(W,N1,N2,N3,REM) of EEG epochs. We used Acknowledgement Software by BIOPAC to complete this process rapidly. Those sub-bands are later used to compute discriminating 75 features.

3.1.3. Extraction of l_1, l_2 and $l\infty$ Norm

The discriminating features used for classifying six different classes (W, S1, S2, S3, S4 and REM) are l_1–norm, $l_2 - norm$ and $l\infty - norm$. The $l_m - norm$ [5] of any discrete-time signal $u[n]$ is defined as

$$\|u\|_m = \left(\sum_{n=1}^{\infty} |u[n]|^m \right)^{\frac{1}{m}}, m \in Z^+$$

$l\infty - norm$ or Peak Absolute Value
The $l\infty - norm$ of a signal gives the maximum absolute value among all the samples

of a discrete-time signal. Thus, it is also known as peak absolute value.

$$\|u\|\infty = |u|_{max}$$

Thus, we get a total of 75 features after combining all three norms.

3.2.1 Feature extraction

Feature extraction is one of the significant step for analysis of sleep behaviour from the EEG signals. The feature-based analysis has been playing a vital role for identifying different sleep characteristics. The brain EEG signals are highly random and non-stationary. Extracting the most relevant features is important for properly interpreting the sleep stages' behaviour. Because The brain EEG signals are not static and constant. It is highly random and non-stationary. Both time and frequency domain features are obtained in this method. Most of the contributions majorly focused less features. Where we obtained 75 features in this method. The individual sleep stages it provides the changes of delta (δ) rhythm (0-4Hz), theta (θ) rhythm(4-8Hz), alpha (α) rhythm(8-12Hz), beta (β) rhythm(12-30Hz), gamma (γ) rhythm(30+Hz) patterns in the EEG signals in a different frequency range. We used three channel F4,C4 and O2. Each channel has five sleep rhythm. Each rhythm has five more sub features. Such as MeanP, MedianF, Spectral Edge, PeakF. So we got total 75 features.

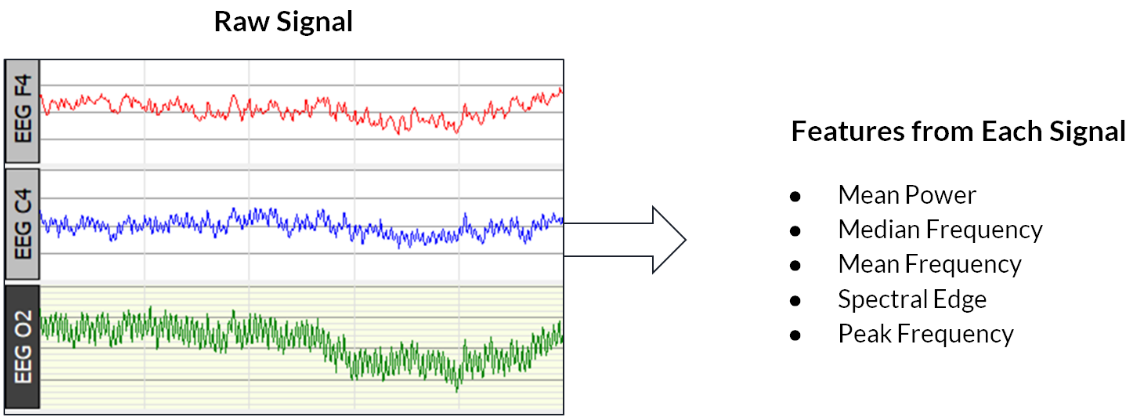

Figure 2: Features

3.2.1 Converting N1,N2,N3,REM to Sleep

In this method, we used sleep and wake data for training segment. Our data set provide us six different sub-bands(W,N1,N2,N3,REM) of EEG epochs. Where N1,N2,N3 are non–rapid eye movement (NREM) and REM is rapid eye movement. W is the wake stage. So we convert NREM and REM into sleep stage and W into wake stage. We got 19355 rows for wake stage and 89097 rows for sleep stage. This two classification is used to train our model to predict this two sleep stages.

Sleep Stage	MeanP_Alpha_F4	MedianF_Alpha_F4	MeanF_Alpha_F4	Spectral Edge_Alpha_F4	PeakF_Alpha_F4
W	0.00051	8.74146	18.48195	17.48293	8.24195
W	0.0004	10.24	17.48293	16.85854	10.61463
W	0.00036	9.74049	17.98244	17.48293	8.74146
W	0.00035	10.11512	17.6078	17.10829	9.61561
W	0.00033	9.74049	17.6078	17.23317	8.11707
W	0.00036	10.24	16.98341	16.35902	10.11512
N1	0.00034	10.11512	17.85756	17.48293	10.24
N1	0.00035	10.73951	17.6078	17.10829	9.1161
N1	0.00035	9.36585	17.6078	16.98341	9.24098
N2	0.0005	10.36488	16.60878	15.85951	7.9922
N2	0.00072	9.36585	15.98439	15.48488	7.9922
N2	0.00069	8.74146	15.23512	14.61073	8.86634

3.3.1 Feature Selection

Every model is highly dependent on the features of the dataset. But large number of features can delay the training phase and the testing phase both. It reduces the complexity of a model and makes it easier to interpret. It improves the accuracy of a model if the right subset is chosen. We also reduced the number of the features in different scale and found out the results. There are 75 features which have been extracted from the data set. But our aim to find the most significant features and the impact of those features on our model. We have used Tree Classifier and Univariate Feature Selection method to find out the most significant features. We segmented 10 features, 15 features and 20 features and all features and trained with that segmented dataset. Univariate Feature Selection method and Tree Classifier both gave same result. But Univariate Feature Selection took less time. So we chose Univariate Feature Selection method for feature selection process.

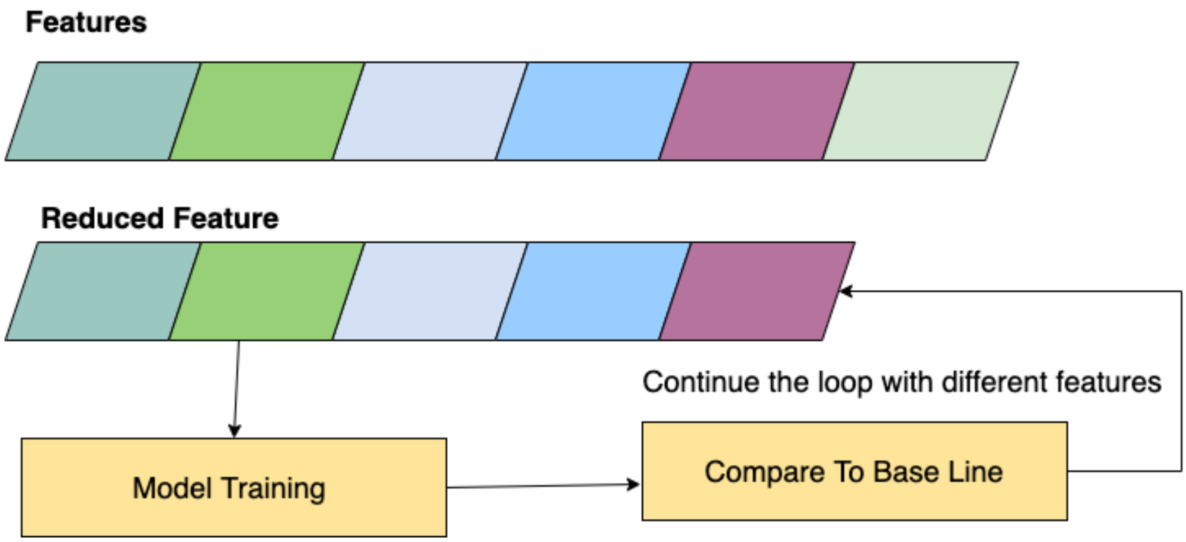

Figure 3: Features Selection

3.3.2 Univariate Feature Selection

In feature-based filter selection, the statistical measures are calculated considering only a single input variable at a time with a target (output) variable. These statistical measures are termed as univariate statistical measures, which means that the interaction between input variables is not considered in the filtering process.[6] This process sets a score to each features according to their importance.We used this selection method and select best 10,15 and 20 features and splitted the data set. The features scores are given bellow.

Features	Scores
Spectral Edge_Beta_F4	11373.488527
MedianF_Beta_F4	9190.314853
Spectral Edge_Beta_C4	8672.075254
Spectral Edge_Alpha_F4	8325.573715
PeakF_Beta_O2	7897.908333
MeanF_Alpha_F4	6863.309690
Spectral Edge_Alpha_C4	5610.763284
MeanF_Beta_F4	5432.923221
MedianF_Beta_C4	4356.787096
MeanF_Alpha_C4	4317.518029

Table 1: Best 10 Features Selection

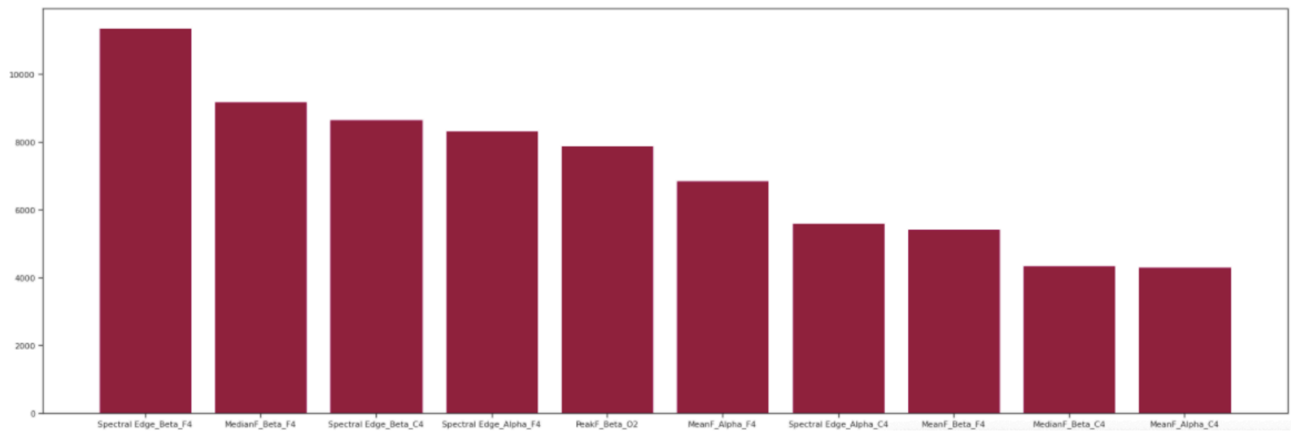

Figure 4: Best10 Features Selection

Features	Scores
Spectral Edge_Beta_F4	11373.488527
MedianF_Beta_F4	9190.314853
Spectral Edge_Beta_C4	8672.075254
Spectral Edge_Alpha_F4	8325.573715
PeakF_Beta_O2	7897.908333
MeanF_Alpha_F4	6863.309690
Spectral Edge_Alpha_C4	5610.763284
MeanF_Beta_F4	5432.923221
MedianF_Beta_C4	4356.787096
MeanF_Alpha_C4	4317.518029
PeakF_Theta_C4	4037.671078
MeanF_Beta_C4	4023.870095
PeakF_Theta_O2	3924.855844
MedianF_Theta_C4	3748.523197
MedianF_Theta_O2	3734.659315

Table 2: Best 15 Features Selection

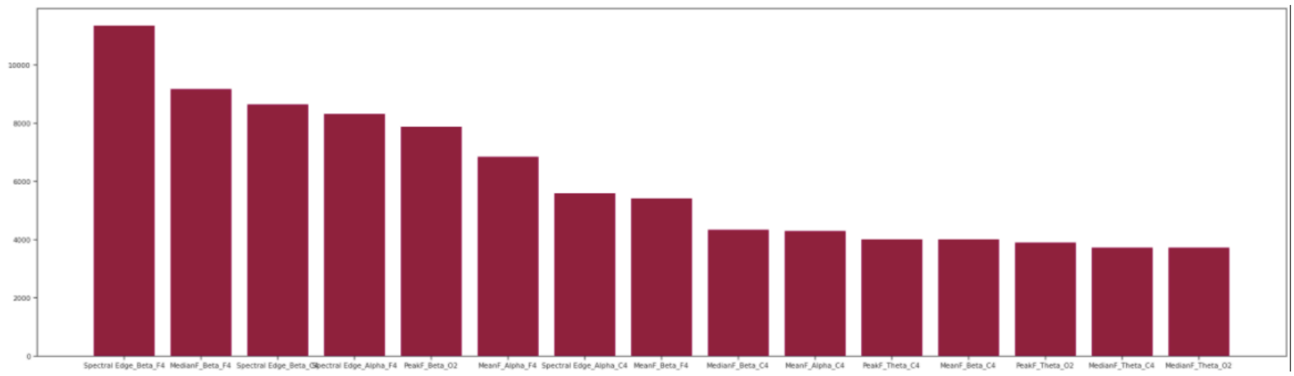

Figure 5: Best15 Features Selection

Features	Scores
Spectral Edge_Beta_F4	11373.488527
MedianF_Beta_F4	9190.314853
Spectral Edge_Beta_C4	8672.075254
Spectral Edge_Alpha_F4	8325.573715
PeakF_Beta_O2	7897.908333
MeanF_Alpha_F4	6863.309690
Spectral Edge_Alpha_C4	5610.763284
MeanF_Beta_F4	5432.923221
MedianF_Beta_C4	4356.787096
MeanF_Alpha_C4	4317.518029
PeakF_Theta_C4	4037.671078
MeanF_Beta_C4	4023.870095
PeakF_Theta_O2	3924.855844
MedianF_Theta_C4	3748.523197
MedianF_Theta_O2	3734.659315
PeakF_Theta_F4	3063.144365
PeakF_Beta_C4	2739.971141
MeanF_Delta_C4	2628.359081
Spectral Edge_Alpha_O2	2596.380657
Spectral Edge_Gamma_F4	2482.252614

Table 3: Best 20 Features Selection

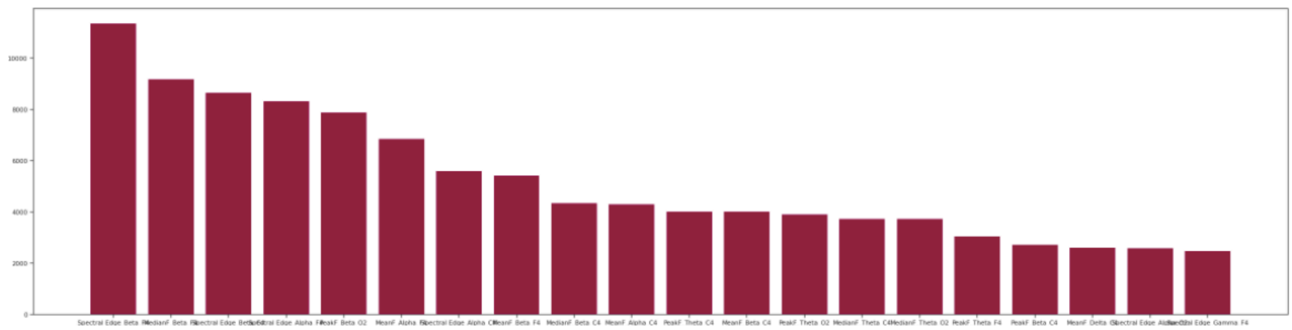

Figure 5: Best 20 Features Selection

3.4.1 Classification

In classification we try several type of Machine Learning algorithm to train our model and compare each type of algorithm's accuracy according to the different segmented data set such as 10 features, 15 feature,20 features and all features. All these segmented features are fed to all the available supervised machine learning classifiers namely TF-DF (TensorFlow Decision Forests), C4.5 algorithm[11], Logistic Regression [7], Naive Bayes [8], Support Vector Machines (SVM) [9], GX Boosted to select the optimum performing classifier.

3.4.1 TensorFlow Decision Forests

TensorFlow Decision Forests (TF-DF) is a collection of state-of-the-art algorithms for the training, serving and interpretation of Decision Forest models. The library is a collection of Keras models and supports classification, regression and ranking.[10] In our proposed method we used 300 tress for each segment. We used the "*tensorflow_decision_forests*" python library to implement the decision forests algorithm.

For 10 features we got 0.8890 accuracy.

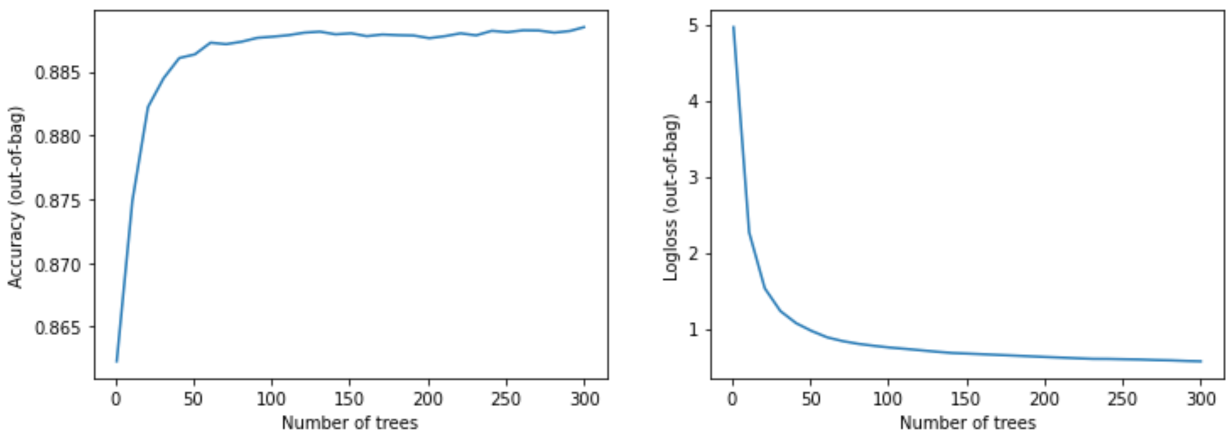

Figure 4: Accuracy graph for 10 features.

For 15 features we got 0.9229 accuracy.

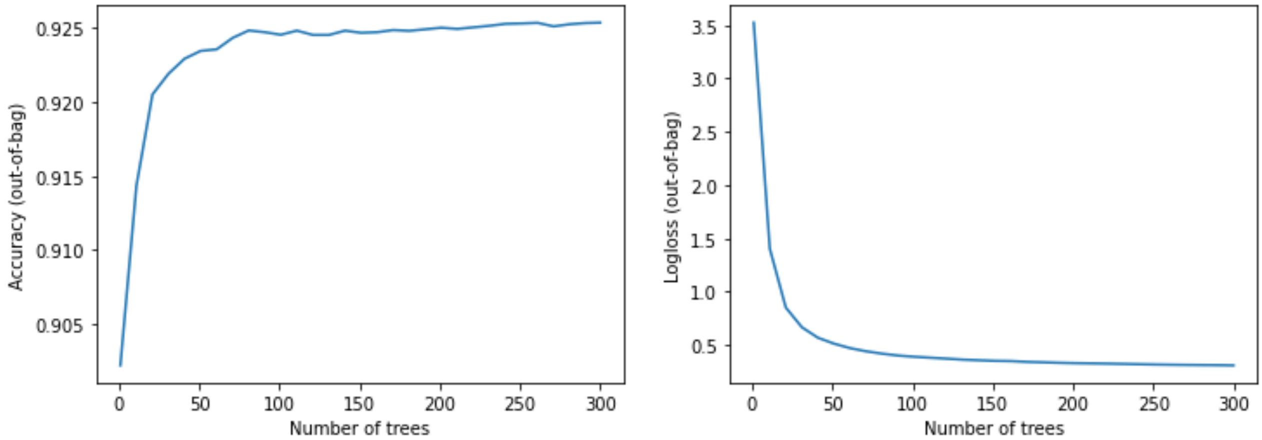

Figure 5: Accuracy graph for 15 features.

For 20 features we got 0.9269 accuracy.

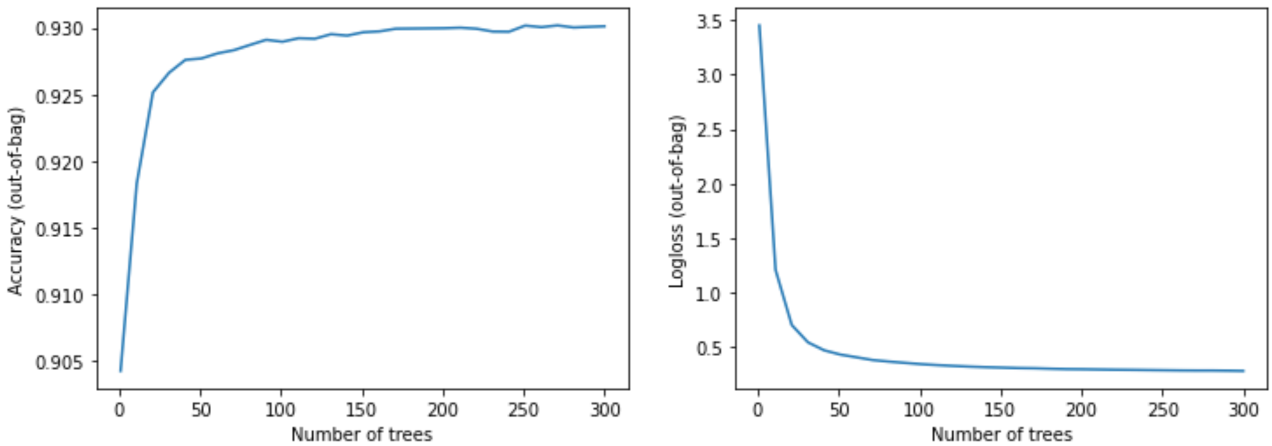

Figure 6: Accuracy graph for 20 features.

For All features we got 0.9413 accuracy.

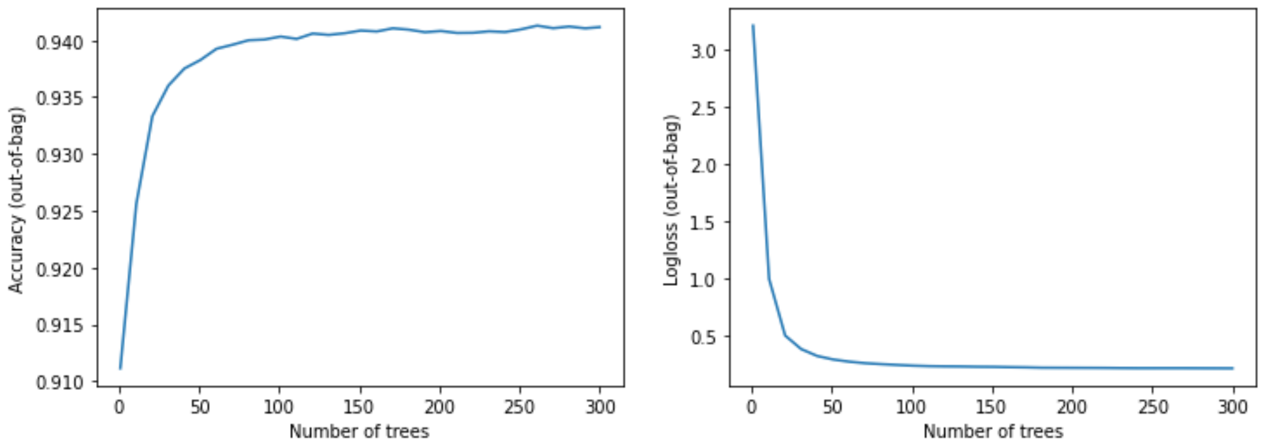

Figure 5: Accuracy graph for all features.

Segments	Accuracy (Total)	Accuracy (10 Best Features)	Accuracy (15 Best Features)	Accuracy (20 Best Features)	Data Set Size (rows)
Sleep/Wake	0.9413	0.889	0.9229	0.9269	108452

Table 4:Accuracy for different segments.

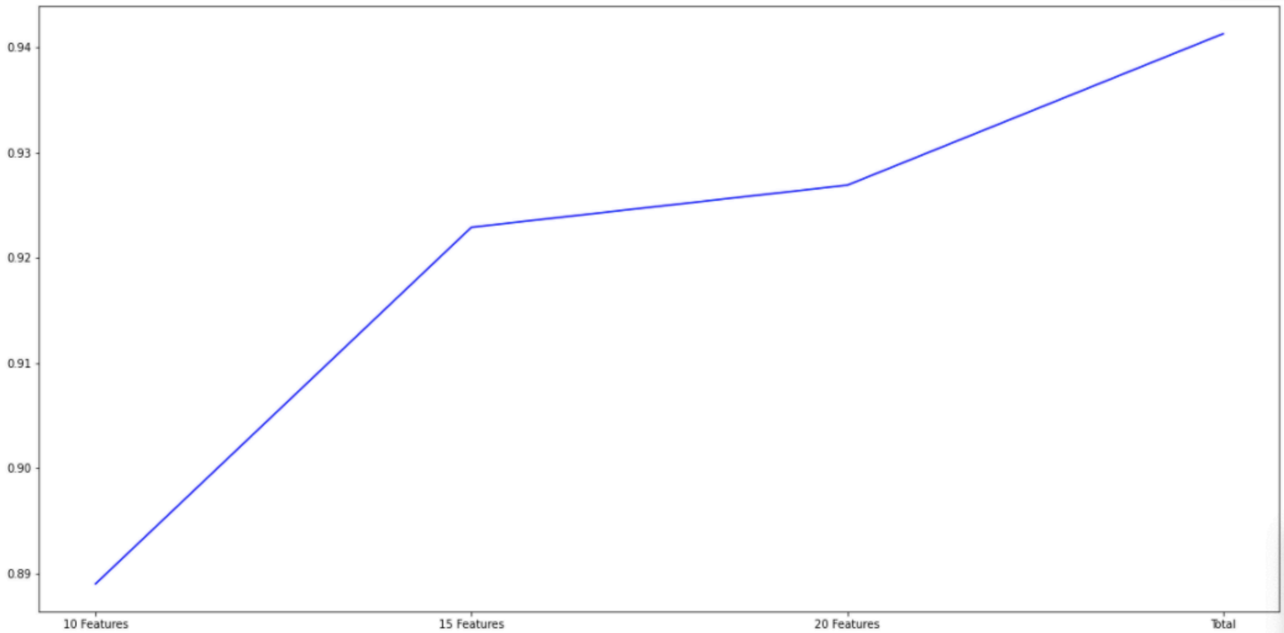

Figure 7: Accuracy graph for all features.

3.4.2 C4.5 Tree Algorithm

C4.5 is a well-known algorithm used to generate a decision trees. C4.5 is a computer program for inducing classification rules in the form of decision trees from a set of given instances. The decision trees generated by the C4.5 algorithm can be used for classification, and for this reason, C4.5 is also referred to as a statistical classifier. At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits data set into subsets that can be one class or the other[11]. The "chefboost" python library has been used for implementing C4.5 tree algorithm.

Segments	Accuracy (Total)	Accuracy (10 Best Features)	Accuracy (15 Best Features)	Accuracy (20 Best Features)	Data Set Size (rows)
Sleep/Wake	0.9248	0.8489	0.9039	0.90489	108452

Table 5:Accuracy for different segments.

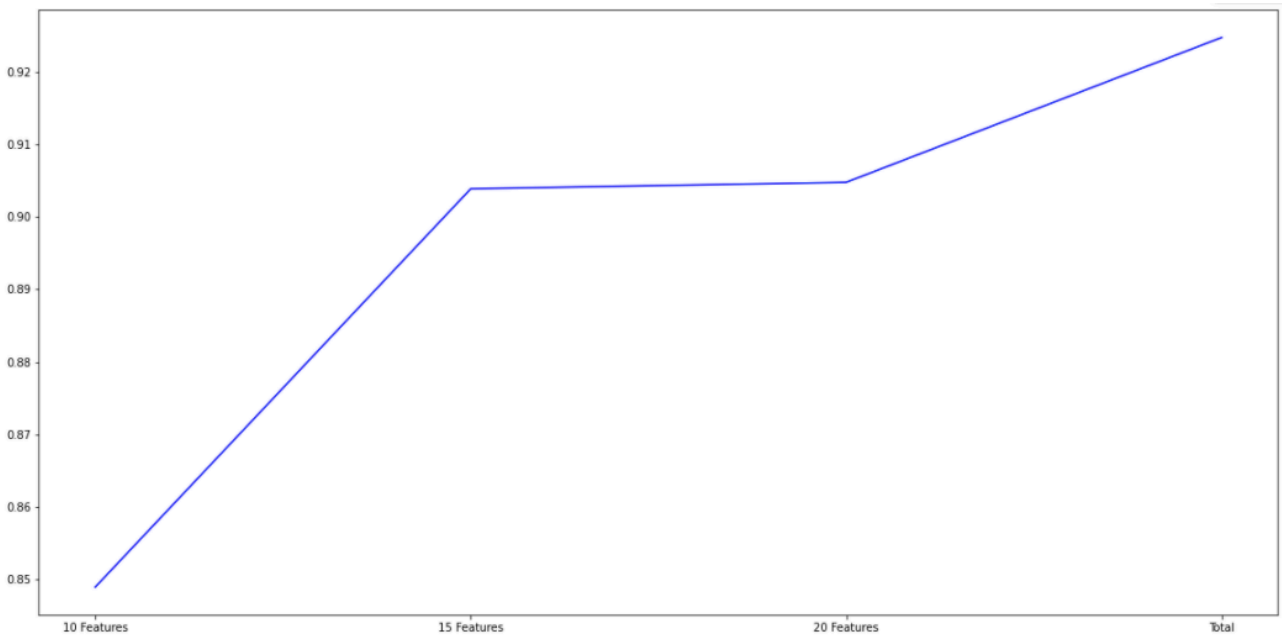

Figure 7: Accuracy graph for all features.

3.4.3 Naive Byes

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable. Bayes' theorem states the following relationship, given class variable y and dependent feature vector x_1 through x_n :

$$P(y \mid x_1, \dots, x_n) = \frac{P(y) P(x_1, \dots, x_n \mid y)}{P(x_1, \dots, x_n)}$$

$$P(y \mid x_1, \dots, x_n) \propto P(y) \prod_{i=1}^{n} P(x_i \mid y)$$

$$\Downarrow$$

$$\hat{y} = \arg\max_{y} P(y) \prod_{i=1}^{n} P(x_i \mid y),$$

We used *"sklearn"*, a python library for implementing Naive Byes. And we got

Segments	Accuracy (Total)	Accuracy (10 Best Features)	Accuracy (15 Best Features)	Accuracy (20 Best Features)	Data Set Size (rows)
Sleep Wake	0.8923	0.8139	0.8684	0.8815	108452

Table 6:Accuracy for different segments.

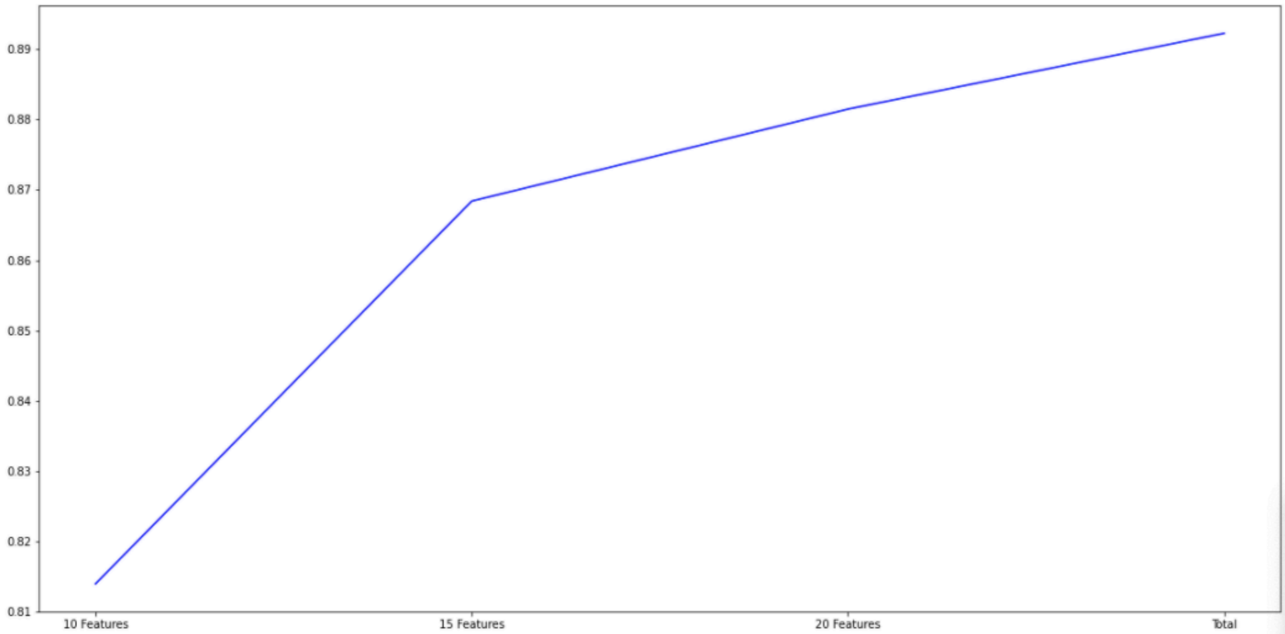

Figure 8: Accuracy graph for all features.

3.4.4 Support vector machines (SVM)

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. The "*sklearn*" python library has been used for implementing SVM.

Segments	Accuracy (Total)	Accuracy (10 Best Features)	Accuracy (15 Best Features)	Accuracy (20 Best Features)	Data Set Size (rows)
Sleep Wake	0.9303	0.8642	0.9148	0.9199	108452

Table 7:Accuracy for different segments.

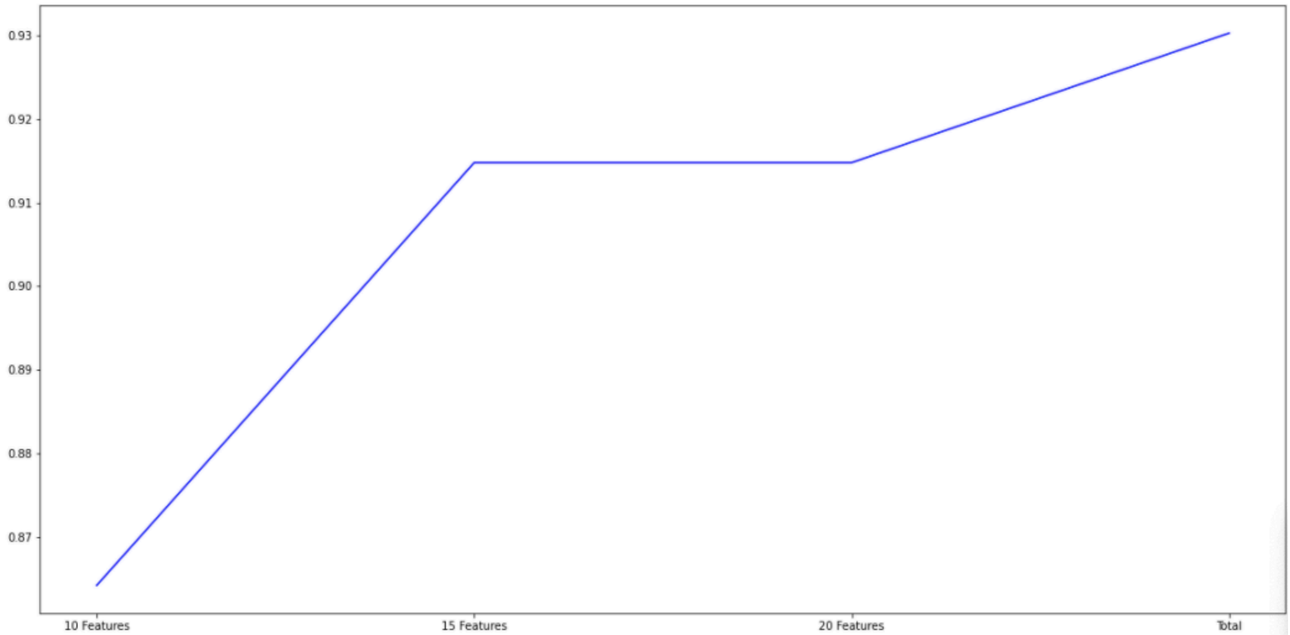

Figure 9: Accuracy graph for all features.

3.4.4 Regression

We used linear regression in this method. Linear regression uses the relationship between the data-points to draw a straight line through all them. Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), consequently called linear regression. We calculated R2 Score for each data segments.

Segments	Accuracy (Total)	Accuracy (10 Best Features)	Accuracy (15 Best Features)	Accuracy (20 Best Features)	Data Set Size (rows)
Sleep Wake	0.9962	0.9963	0.9984	0.9922	108452

Table 8:Accuracy for different segments.

{to be continued}

3.5.1 Prediction Result

We got different accuracy result using different algorithms. We used 10,15, 20 and all features to train the model in various approach. We can observe that, TensoreFlow Decision Forest algorithm has performed better than others. It hold its performance both in less and data.

3.5.2 Prediction Result for 10 features

We can see in figure 10, The TF-DF (TensoreFlow Decision Forest) has 0.889 accuracy for 10 features. Which significantly dropped in C4.5 algorithm (0.8489). SVM and Naive Byes have

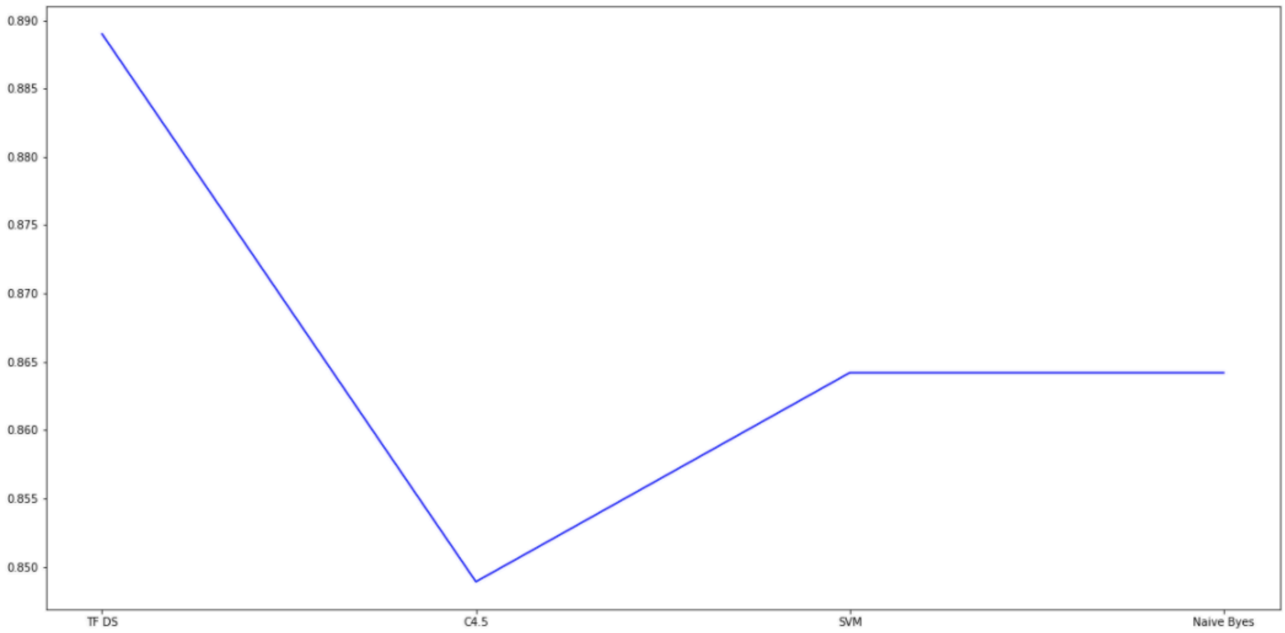

Figure 10: Accuracy graph for 10 features

performed almost equally. In SVM we got 0.8642 and 0.8684 in Naive Byes.

3.5.3 Prediction Result for 15 features

The TF-DF holds its performance with 15 features. Which is 0.9229, slightly higher than the 10 features data set. In this segmentation, C4.5 has performed well. It showed 0.9039 accuracy, which is better than the previous segmentation. The accuracy of SVM has increased and showed 0.9148.

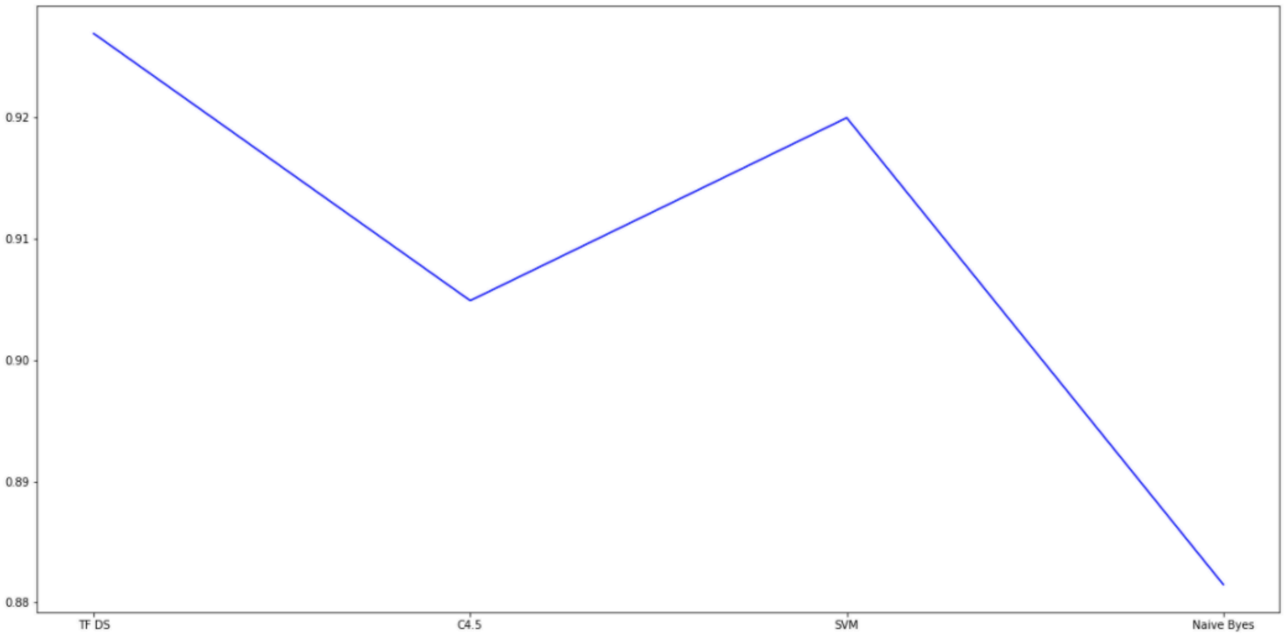

Figure 11: Accuracy graph for 15 features

Naive Byes disappointed in this segmentation. Though the accuracy increased according to the number of features increment, it is far behind from the other other algorithms. It showed 0.8684 accuracy.

3.5.4 Prediction Result for 20 features

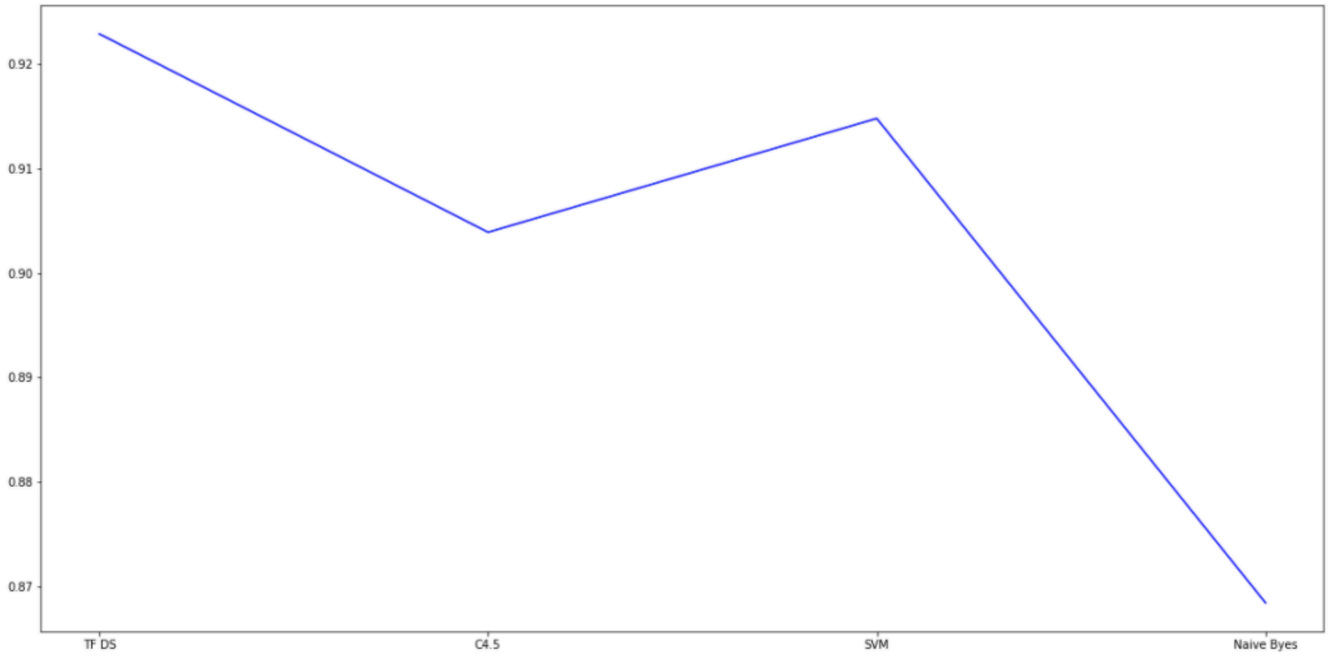

Figure 12: Accuracy graph for 20 features

This segmentation showed almost same pattern in graph as previous. The highest score was 0.9269 by the TF-DF. Naive Byes increased its accuracy but it was far behind from the other algorithms. It scored 0.8815. C4.5 scored 0.9048 and SVM scored 0.9196. The accuracy differences between this two algorithms slightly decreased.

3.5.5 Prediction Result for All features

When we considered the total 75 features TF-DF scored highest again. It scored 0.9413. SVM hold

Figure 12: Accuracy graph for 20 features

the second position. It scored 0.9303. Naive Byes fall behind once more time. It fall behind from each algorithms with accuracy score 0.8923. The C4.5 scored 0.9248. The accuracy differences between C4.5 and SVM again decreased.

3.5.6 Overview

We can observe that TF-DF (TensoreFlow Decision Forest) showed a great performance among the algorithms.

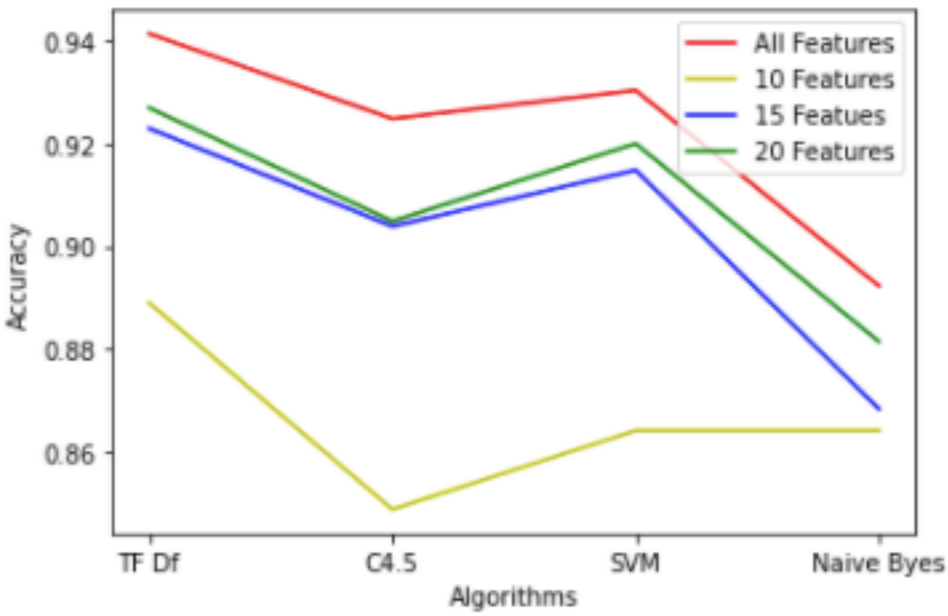

Figure 13: Combined Graph

[1] Polysomnography and Sleep Disorders, Jean K. Matheson, Randip Singh, and Andreja Packard,
[2]Aboalayon Khald, Ocbagabir Helen, Faezipour, Miad. Efficient Sleep Stage Classification Based on EEG Signals.2014 IEEE Long Island Systems, Applications and Technology Conference, LISAT 2014.
[3]Machine learning with ensemble stacking model for automated sleep staging using dual-channel EEG signal, Santosh Kumar Satapathy , Akash Kumar Bhoi, D. Loganathan, Bidita Khandelwal, Paolo Barsocchi,Biomedical Signal Processing and Control 69 (2021) 102898
[4] Obayya Marwa, abou-chadi, Fatma. Automatic classification of sleep stages using EEG records based on Fuzzy c-means (FCM) algorithm. National Radio Science Conference, NRSC, Proceedings;2014, p. 265-272.
[5] Strang, G.; Strang, G.; Strang, G.; Strang, G. Introduction to Linear Algebra; Wellesley-Cambridge Press: Wellesley, MA, USA, 1993; Volume 3.
[6] POPULAR FEATURE SELECTION METHODS IN MACHINE LEARNING,Anber Arif Dataaspirant,December 28, 2020

*[7] Tolles, J.; Meurer, W.J. Logistic regression: Relating patient characteristics to outcomes. JAMA **2016**, 316, 533–534.*

[8]Lewis, D.D. Naive (Bayes) at forty: The independence assumption in information retrieval. In European Conference on Machine
Learning; Springer: Berlin/Heidelberg, Germany, 1998; pp. 4–15.

*[9]Cortes, C.; Vapnik, V. Support-vector networks. Mach. Learn. **1995**, 20, 273–297*

[10] TensorFlow Decision Forests, https://www.tensorflow.org/decision_forests.
[11] PREDICTING STUDENTS' PERFORMANCE USING ID3 AND C4.5 CLASSIFICATION ALGORITHMS, Kalpesh Adhatrao, Aditya Gaykar, Amiraj Dhawan, Rohit Jha and Vipul Honrao,International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.3, No.5, September 2013