Title Page:

Accurate Detection of Human Stress with Heart Rate and Blood Pressure Data for Elderly Patients using KNN Algorithm in Comparison with Naive Bayes

Gundlooru Supriya[1], Dr. Kirupa Ganapathy[2]

Gundlooru Supriya[1]

Research Scholar,

Department of Electronics and Communication Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamilnadu. India. Pincode: 602105.

gundloorusupriya17@saveetha.com

Prof. Kirupa Ganapathy[2]

Project Guide, Corresponding Author,

Department of Electronics and Communication Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamilnadu. India. Pincode: 602105.

kirupagaapathy.sse@saveetha.com

Keywords: K Nearest Neighbor, Novel Naive Bayes, Mental Stress, Machine Learning, Kaggle, Heart Rate, Blood Pressure.

ABSTRACT

Aim: This study aims to detect mental stress in elderly patients using the K Nearest Neighbor (KNN) classifier. To detect mental stress, heart rate and blood pressure data is used. **Materials and Methods:** A total of 41034 samples are collected from a dataset available in kaggle (SWELL dataset). These samples are divided into training dataset (n = 28724 [70%]) and test dataset (n = 12310 [30%]). Accuracy is calculated to quantify the performance of the KNN algorithm. **Results:** The KNN algorithm achieved an accuracy of 91.3%, and 87% whereas Naive Bayes achieved an accuracy of 88%, and 83% ($P<0.05$) in SWELL and Biometrics for stress monitoring datasets respectively. The G power is calculated as 0.8. **Conclusion:** In this study, it is observed that the KNN algorithm performs significantly better than the Naive Bayes algorithm.

Keywords: K Nearest Neighbor, Novel Naive Bayes, Mental Stress, Machine Learning, Kaggle, Heart Rate, Blood Pressure.

INTRODUCTION

In these fast-paced times, psychological health issues like stress, anxiety, and depression became quite common among people (Priya, Garg, and Tigga 2020). The symptoms of stress are feeling upset or agitated, an inability to relax, low energy levels, chronic headaches, frequent overreaction, and persistent colds or infections. Due to mental health issues, one-third of the population in the world undergoes depression, suicide (Shafiee et al. 2020). In this paper, the proposed work gives classification and detection of mental stress in the elderly using machine learning algorithms like KNN (K Nearest Neighbor) in comparison with Naive Bayes (P. Kumar, Garg, and Garg 2020). An increase in mental stress in people leads to major problems like hypertension, anxiety, chronic illness (Can et al. 2019a). The proposed work can be applied to students, employed and unemployed individuals, and elderly people (Ahuja and Banga 2019).

Several research articles were published on KNN and Naive Bayes in the past five years. 1090 articles were published in google scholar, and 210 articles were published on the web of science. The stress of a person is obtained by heart rate variability which is calculated from the PPG signal (Nkurikiyeyezu et al. 2019) but the PPG signal only measures the rate of blood flow as controlled by the heart's pumping action so the efficiency is less. Stress is a broad concept referring to psychological and biological processes during emotional and cognitive demanding situations (Koldijk, Neerincx, and Kraaij 2018). This paper used the ECG, EEG, and GSR signals for detecting the stress in the individuals but these signals took time for frequency and spectral analysis. Stress reduces human functionality during routine work and may lead to severe health defects (Attallah 2020). But this paper detects the stress using ECG, EEG signals so the signal analysis takes a lot of time. By using statistical parametric analysis from the time domain, and wavelet-based bandwidth specific feature analysis from the time-frequency domain EEG signals can be processed to detect the mental stress of an individual (Hasan and Kim 2019). Stress can be experienced daily from a variety of different reasons, including environmental reasons (traffic, noise, or bad weather), social reasons (family issues, friends, and financial problems) (Elzeiny and Qaraqe 2018). The mental stress detection was done using the recurrence quantification analysis method, which means they used heart rate and respiratory signals to analyze and pre-proceed with machine learning algorithms (Fernández and Anishchenko 2018).

But these analyses consume more time and are also difficult to use on elderly persons. Using EEG signals the mental stress detected with the KNN algorithm (Sha'abani et al. 2020) but in this paper, they used only EEG signals for analyzing and detecting so the efficiency is less. Stress classification was extracted by using ECG and heart rate variability signals with different machine learning algorithms (Dalmeida and Masala, n.d.) but they used a wearable device for classification, so first,they needed to analyze the ECG signals which consume a lot of time. Stress in working people was detected using the ECG and GSR signals by using KNN and other various machine learning algorithms which is difficult to use for older individuals (Sriramprakash, Prasanna, and Ramana Murthy 2017). Mental stress was classified and detected by using EEG signals at five different features which include power spectral density, correlation, differential asymmetry, rational asymmetry, and power spectrum. Stress was detected by using different processing methods of EEG signals like EEG feature extraction, feature selection which includes receiver operating characteristic curve, t-test and the Bhattacharya distance, classification done by different machine learning algorithms, and tenfold cross-validation.

Previously our team had rich experience in working on various research projects across multiple disciplines.(M. S. Kumar et al. 2006; Mehta et al. 2019; Neelakantan et al. 2011; Praveen et al. 2001) . Now the growing trend in this area motivated us to pursue this project.

The existing works predicted and detected mental stress using various frequency and spectral signals which are generated from physiological sensors like GSR (Galvanic Skin Response), ECG (Electrocardiogram), and EEG (electroencephalogram). The classification and detection of mental stress using these methods were difficult and they consumed a lot of time for analysis. In this proposed work, the KNN is used to detect mental stress in elderly patients and the performance of the proposed technique is compared with novel Naive Bayes.

MATERIALS AND METHODS

The proposed work is carried out in the machine learning laboratory of Saveetha School of Engineering. The proposed work consists of a group for mental stress detection. The total sample size of the group is 41034 samples. The minimum number of samples required for this study is calculated in clincalc calculator with G power = 0.8, the maximum error rate is fixed as 0.2 and alpha = 0.05 (Novani et al. 2018).

The SWELL dataset is publicly available on Kaggle website which comprises heart rate variability (HRV) indices computed from the multimodal SWELL knowledge work (SWELL-KW) dataset for research on stress and user modeling. The SWELL was collected by researchers at the Institute for Computing and Information Sciences at Radboud University. It is a result of experiments conducted on 25 subjects doing typical office work (for example writing reports, making presentations, reading email, and searching for information) (qiriro n.d.). The subject went through typical working stressors such as receiving unexpected email interruptions and pressure to complete their work on time. The experiment recorded various data including computer logging, facial expression, body postures, ECG signal, and skin conductance. The researchers also recorded the subjective experience on task load, mental effort, emotion, and perceived stress. Each participant went through three different working conditions like no stress, time pressure, interruption.

The Biometrics for stress monitoring dataset is available on the Kaggle website which comprises heart rate variability (HRV) and Electrodermal activity (EDA) features. The Biometrics for stress monitoring dataset was collected from TNO & Radboud University. It contains 16 different attributes for classification and detecting mental stress in humans.

The processed dataset is given for both training and testing. Data processing includes missing data removal, replacement of null values and float values with mean or median values, and standardization of data. The preprocessed dataset with features is given as input to KNN and Naive Bayes. From the total sample size, 70% of the data is given for training and the remaining 30% is given for testing. Both the datasets are used in this proposed work for group1 and group2 binary classification as given in Table 1.

Table 1 represents the sample features and classes of the input dataset for analysis. The SWELL dataset contains 12 features with 2 different classes and the Biometrics for stress monitoring dataset contains 16 attributes with 2 classes.

KNN Algorithm

KNN is a supervised machine learning algorithm used for solving regression and classification problems. It is a simple and non-parametric algorithm. The working of KNN is as follows:

- Load the data.
- Initialize K to the chosen number of neighbors.
- For each data in the dataset.
- Calculate the distance between the query data and the current data from the dataset.
- Add the distance and the index of the data to an ordered collection.
- Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances.
- Pick the first K entries from the sorted collection.
- Get the labels of the selected K entries.
- If regression, return the mean of the K labels.
- If classification, return the mode of the K labels.

Naive Bayes Algorithm

Naive Bayes is a classification algorithm for binary (two-class) and multiclass classification problems. Novel Naive Bayes or idiot Bayes is used to calculate the probabilities for each class which is simplified to make their calculations tractable. The probability of the classes is calculated by the equation:

$$\mathbf{P(class|data) = (P(data|class) * P(class)) / P(data)} \quad \mathbf{(1)}$$

Where P(class|data) is the probability of class to the provided data. The working of KNN is:

- Separate into Class.
- Summarize Dataset.
- Summarize Data by Class.
- Gaussian Probability Density Function.
- Class Probabilities.

The test setup used for this classification is a core i5 processor with 4.00GB RAM. All the simulations are performed in Google Colabs. From the total sample size, 70% of the data is given for training and the remaining 30% is given for testing. The accuracy can be calculated using true positive, true negative, false positive, and false negative values.

All the statistical analysis is conducted using SPSS and python tools. Descriptive statistics (mean, standard deviation, and standard error) is carried out for KNN and Naive Bayes algorithms. Independent variables in this study are the input variables (Age, Sex, Education, cholestrol, Systolic blood pressure, Dystolic blood pressure, Skin temperature, Skew, Kurt, BMI, Medication, Current Smoker, Heart Rate, Hyp, diabetes, Stressed). The dependent variable is the output variable (Accuracy). An Independent t-test is performed to compare the performance of KNN and Naive Bayes.

RESULTS

Table 1 represents the sample features and classes of the input dataset for analysis. The SWELL dataset contains 12 features with 2 different classes and the Biometrics for stress monitoring dataset contains 16 attributes with 2 classes.

Table 2 shows the comparison of KNN and Naive Bayes algorithms on SWELL and Biometrics for stress monitoring datasets. KNN achieved an accuracy of 91.3% and Naive Bayes achieved an accuracy of 88% in the SWELL dataset. Similarly in the Biometrics for stress monitoring dataset the KNN achieved an accuracy of 87% and the Naive Bayes achieved an accuracy of 83%.

From Fig. 1, it is observed that the accuracy increases as the number of iterations increases. The accuracy value becomes constant after the 300th iteration. We took 10 different iteration samples for checking the accuracy values for the algorithm.

Table 3 shows the analysis of KNN and Naive Bayes algorithms in terms of mean accuracy, standard deviation, standard error mean values. The standard deviation of KNN is 0.01273 which shows higher performance than the Naive Bayes which had standard deviation as 0.01384.

Table 4 shows the mean differences, standard error differences, and significance values between the two groups. For the proposed work the significance value is 0.032 ($p<0.05$).

Figure 2 shows the comparison of KNN and Naive Bayes algorithms as the number of iterations varied. As the iterations vary the accuracy of the algorithm varies, after the 300th iteration the accuracy remains constant.

Figure 3 shows the comparison of accuracy values of KNN and Naive Bayes algorithms as a bar chart by using the SPSS tool. Where the mean accuracy of KNN is higher than the Naive Bayes. The standard deviation of KNN is lesser than the Naive Bayes so, KNN performs better than Naive Bayes.

DISCUSSIONS

In this study, we observed that KNN performs classification better than Naive Bayes with an accuracy of 91.3%, and 88% in the SWELL dataset and Biometrics for stress monitoring dataset in comparison with the accuracy of 87%, and 83% Naive Bayes ($p<0.05$).

In this study, the KNN and Naive Bayes algorithms detected mental stress. The results show that The KNN performs significantly better than the Naive Bayes. Some of the previous studies reported an accuracy of 80.39% for mental stress detection (Can et al. 2019b) which is less accurate compared to the proposed method. In this paper reported that the accuracy of 82% depends on the correlation of the algorithms which is less than the proposed work. These results are more similar to the proposed work. The accuracy values obtained from the two dataset shows that the KNN is more stable compared to the Naive Bayes since the variation in accuracy value for the KNN algorithm for the various number of iterations is less compared to the Naive Bayes algorithm.

Our institution is passionate about high-quality evidence-based research and has excelled in various fields (Vijayashree Priyadharsini 2019; Ezhilarasan, Apoorva, and Ashok Vardhan 2019; Ramesh et al. 2018; Mathew et al. 2020; Sridharan et al. 2019; Pc, Marimuthu, and Devadoss 2018; Ramadurai et al. 2019). We hope this study adds to this rich legacy.

The limitations for this proposed work is KNN becomes significantly slower as the size of the data is increased. It is not efficient when the dataset is very large. In the future, KNN accuracy can be increased with more parameters and large data in training sets. Using real-time data hypertension, coronary heart disease, cardiovascular diseases, and chronic illness can also be prevented.

CONCLUSION

From this proposed work, one can perform an initial analysis for mental stress detection to help the individuals to undergo treatment at the earliest. It is also observed that the proposed KNN algorithm performed significantly better than the Naive Bayes algorithm in mental stress detection. There is an improvement in the accuracy of the proposed algorithm.

DECLARATIONS

Conflict of interests

No conflict of interest in this manuscript.

Authors Contribution

Author GS was involved in data collection, data analysis, manuscript writing. Author KG was involved in conceptualization, data validation, and critical review of the manuscript.

Acknowledgment

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences(Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding: We thank the following organizations for providing financial support that enabled us to complete the study.

1. Biozone Research Technologies Pvt Ltd, India.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

Ahuja, Ravinder, and Alisha Banga. 2019. “Mental Stress Detection in University Students Using Machine Learning Algorithms.” *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2019.05.007.

Attallah, Omneya. 2020. “An Effective Mental Stress State Detection and Evaluation System Using Minimum Number of Frontal Brain Electrodes.” *Diagnostics (Basel, Switzerland)* 10 (5). https://doi.org/10.3390/diagnostics10050292.

Can, Yekta Said, Niaz Chalabianloo, Deniz Ekiz, and Cem Ersoy. 2019a. “Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study.” *Sensors*. https://doi.org/10.3390/s19081849.

Dalmeida, Kayisan Mary, and Giovanni Luca Masala. n.d. “Stress Classification of ECG-Derived HRV Features Extracted from Wearable Devices.” https://doi.org/10.20944/preprints202103.0644.v1.

Elzeiny, Sami, and Marwa Qaraqe. 2018. “Machine Learning Approaches to Automatic Stress Detection: A Review.” *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*. https://doi.org/10.1109/aiccsa.2018.8612825.

Ezhilarasan, Devaraj, Velluru S. Apoorva, and Nandhigam Ashok Vardhan. 2019. “Syzygium Cumini Extract Induced Reactive Oxygen Species-Mediated Apoptosis in Human Oral Squamous Carcinoma Cells.” *Journal of Oral Pathology & Medicine: Official Publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology* 48 (2): 115–21.

Fernández, José Raúl Machado, and Lesya Anishchenko. 2018. “Mental Stress Detection Using Bioradar Respiratory Signals.” *Biomedical Signal Processing and Control*. https://doi.org/10.1016/j.bspc.2018.03.006.

Hasan, Md Junayed, and Jong-Myon Kim. 2019. “A Hybrid Feature Pool-Based Emotional Stress State Detection Algorithm Using EEG Signals.” *Brain Sciences* 9 (12). https://doi.org/10.3390/brainsci9120376.

Koldijk, Saskia, Mark A. Neerincx, and Wessel Kraaij. 2018. “Detecting Work Stress in Offices by Combining Unobtrusive Sensors.” *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/taffc.2016.2610975.

Kumar, Muthusamy Senthil, Gelli Vamsi, Ramasamy Sripriya, and Praveen Kumar Sehgal. 2006. “Expression of Matrix Metalloproteinases (MMP-8 and -9) in Chronic Periodontitis Patients with and without Diabetes Mellitus.” *Journal of Periodontology* 77 (11): 1803–8.

Kumar, Prince, Shruti Garg, and Ashwani Garg. 2020. “Assessment of Anxiety, Depression and Stress Using Machine Learning Models.” *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2020.04.213.

Mathew, M. G., S. R. Samuel, A. J. Soni, and K. B. Roopa. 2020. “Evaluation of Adhesion of Streptococcus Mutans, Plaque Accumulation on Zirconia and Stainless Steel Crowns, and Surrounding Gingival Inflammation in Primary ….” *Clinical Oral Investigations*. https://link.springer.com/article/10.1007/s00784-020-03204-9.

Mehta, Meenu, Deeksha, Devesh Tewari, Gaurav Gupta, Rajendra Awasthi, Harjeet Singh, Parijat Pandey, et al. 2019. “Oligonucleotide Therapy: An Emerging Focus Area for Drug Delivery in Chronic Inflammatory Respiratory Diseases.” *Chemico-Biological Interactions* 308 (August): 206–15.

Neelakantan, P., C. Subbarao, C. V. Subbarao, G. De-Deus, and M. Zehnder. 2011. “The Impact of Root Dentine Conditioning on Sealing Ability and Push-out Bond Strength of an Epoxy Resin Root Canal Sealer." *International Endodontic Journal* 44 (6): 491–98.

Nkurikiyeyezu, Kizito, Kana Shoji, Anna Yokokubo, and Guillaume Lopez. 2019. "Thermal Comfort and Stress Recognition in Office Environment." *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies*. https://doi.org/10.5220/0007368802560263.

Novani, Nefy Puteri, Lathifah Arief, Rima Anjasmara, and Ary Setijadi Prihatmanto. 2018. "Heart Rate Variability Frequency Domain for Detection of Mental Stress Using Support Vector Machine." *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*. https://doi.org/10.1109/icitsi.2018.8695938.

Pc, J., T. Marimuthu, and P. Devadoss. 2018. "Prevalence and Measurement of Anterior Loop of the Mandibular Canal Using CBCT: A Cross Sectional Study." *Clinical Implant Dentistry and Related Research*. https://europepmc.org/article/med/29624863.

Praveen, K., V. Narayanan, M. R. Muthusekhar, and M. F. Baig. 2001. "Hypotensive Anaesthesia and Blood Loss in Orthognathic Surgery: A Clinical Study." *The British Journal of Oral & Maxillofacial Surgery* 39 (2): 138–40.

Priya, Anu, Shruti Garg, and Neha Prerna Tigga. 2020. "Predicting Anxiety, Depression and Stress in Modern Life Using Machine Learning Algorithms." *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2020.03.442.

qiriro. n.d. "SWELL Dataset." Accessed April 15, 2021. https://www.kaggle.com/qiriro/swell-heart-rate-variability-hrv.

Ramadurai, Neeraja, Deepa Gurunathan, A. Victor Samuel, Emg Subramanian, and Steven J. L. Rodrigues. 2019. "Effectiveness of 2% Articaine as an Anesthetic Agent in Children: Randomized Controlled Trial." *Clinical Oral Investigations* 23 (9): 3543–50.

Ramesh, Asha, Sheeja Varghese, Nadathur D. Jayakumar, and Sankari Malaiappan. 2018. "Comparative Estimation of Sulfiredoxin Levels between Chronic Periodontitis and Healthy Patients - A Case-Control Study." *Journal of Periodontology* 89 (10): 1241–48.

Sha'abani, M. N. A. H., M. N. A. Sha'abani, N. Fuad, Norezmi Jamal, and M. F. Ismail. 2020. "kNN and SVM Classification for EEG: A Review." *Lecture Notes in Electrical Engineering*. https://doi.org/10.1007/978-981-15-2317-5_47.

Shafiee, Nor Safika Mohd, Faculty of Computer & Mathematical Sciences, Universiti Teknologi Mara, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia, Sofianita Mutalib, et al. 2020. "Prediction of Mental Health Problems among Higher Education Student Using Machine Learning." *International Journal of Education and Management Engineering*. https://doi.org/10.5815/ijeme.2020.06.01.

Sridharan, Gokul, Pratibha Ramani, Sangeeta Patankar, and Rajagopalan Vijayaraghavan. 2019. "Evaluation of Salivary Metabolomics in Oral Leukoplakia and Oral Squamous Cell Carcinoma." *Journal of Oral Pathology & Medicine: Official Publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology* 48 (4): 299–306.

Sriramprakash, S., Vadana D. Prasanna, and O. V. Ramana Murthy. 2017. "Stress Detection in Working People." *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2017.09.090.

Vijayashree Priyadharsini, Jayaseelan. 2019. "In Silico Validation of the Non-Antibiotic Drugs Acetaminophen and Ibuprofen as Antibacterial Agents against Red Complex Pathogens." *Journal of Periodontology* 90 (12): 1441–48.

Tables and Figures

Table 1: Input data for analysis. Both SWELL and Mental Stress are suitable for binary classification.

Dataset	No.of Patients	Features	Classes
SWELL	32794	12	2
Biometrics for stress monitoring	8240	16	2

Table 2: Comparison of KNN and Naive Bayes with accuracy.

Datasets	KNN	Naive Bayes
SWELL dataset	Accuracy - 91.3%	Accuracy - 88%
Biometrics for stress monitoring dataset	Accuracy - 87%	Accuracy - 83%

Table 3: Represents the performance of KNN and Naive Bayes algorithms in terms of mean accuracy, standard deviation and standard error mean. The mean accuracy of KNN is higher than Naive Bayes.

	Groups	**N**	**Mean**	**Std. Deviation**	**Std.Error Mean**
ACCURACY	KNN	10	0.8625	0.01273	0.00592
	Naive Bayes	10	0.8339	0.01384	0.00438

Table 4: Mean differences, Standard error differences and significance values calculated from the independent t sample test. The significance for the proposed work is 0.032 ($P<0.05$).

Leven's Test For Equality of Variance				**t-test for Equality of Variance**					**95% Confidence Interval of the difference**	
		F	sig.	t	dif	sig(2-tailed)	Mean difference	Std.Error Difference	lower	upper
	Equal Variance assumed	0.008	0.032	3.3888	18	.001	0.02863	0.00736	.01306	.04410
Accuracy	Equal variance not assumed			3.3888	16.568	.001	0.02863	0.00736	.01306	.04420

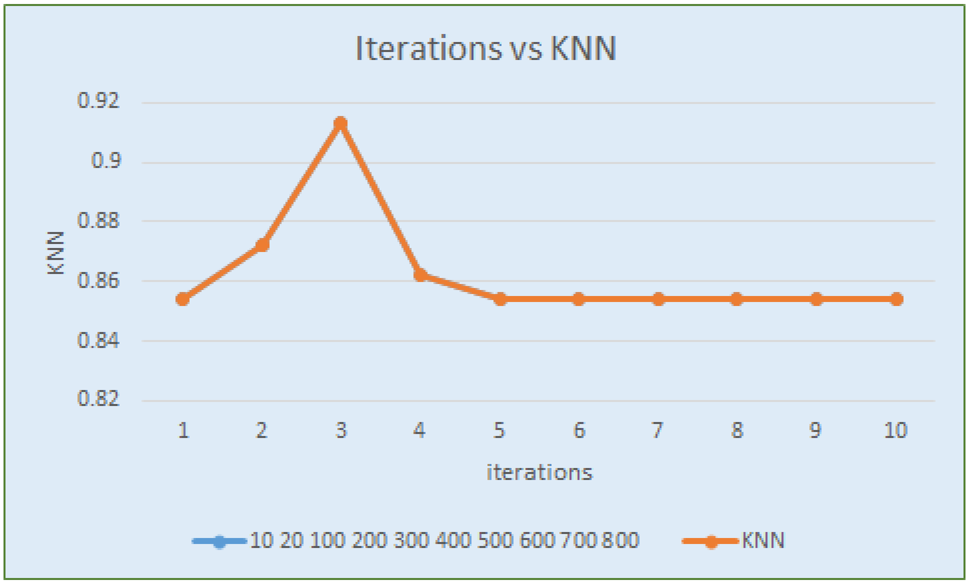

Fig.1. Accuracy of KNN for different iterations. Fluctuations occur before the 300th iteration after that the accuracy remains constant.

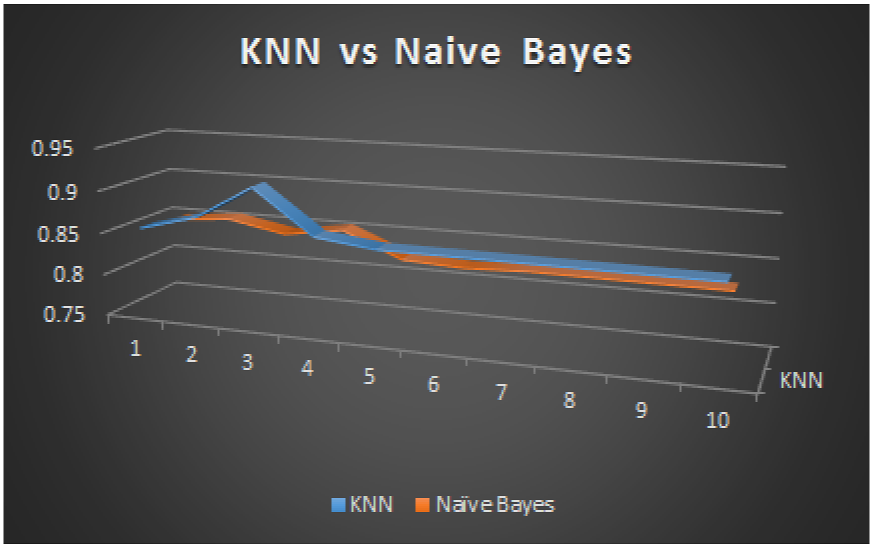

Fig.2. Comparison of KNN and Naive Bayes at different iteration values

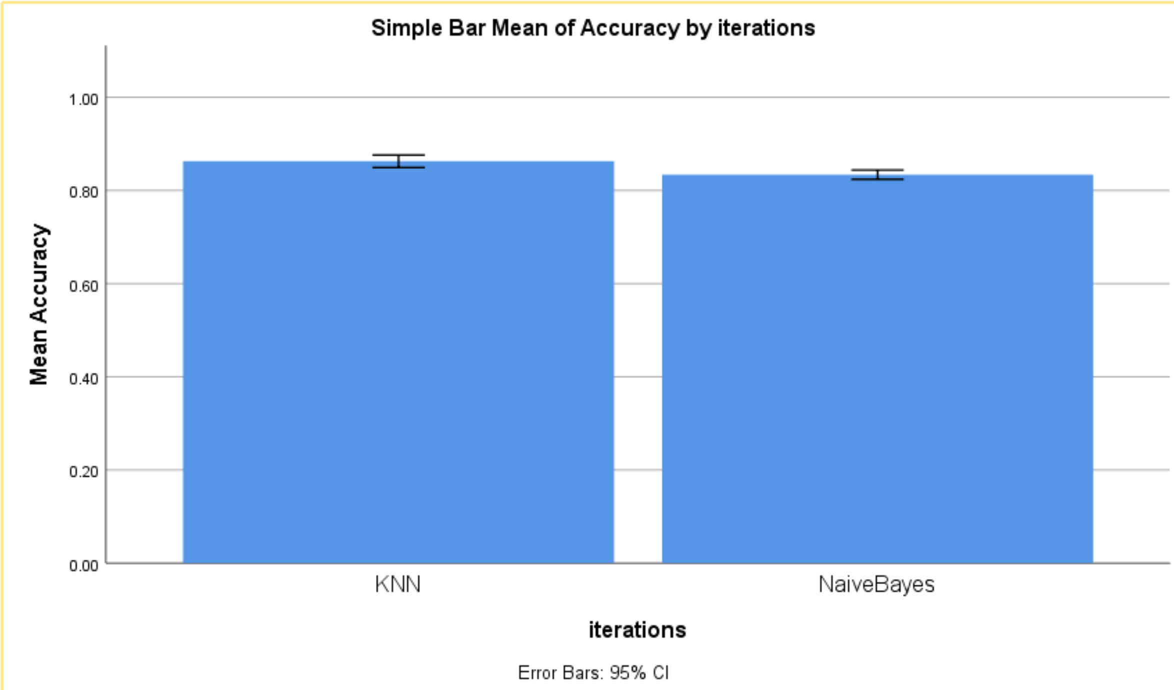

Fig.3. In terms of mean accuracy, the KNN and Naive Bayes are compared. The X-axis represents the groups (algorithms) and Y-axis represents the mean accuracy. It shows that the KNN mean accuracy is higher than the Naive Bayes and the Mean accuracy is ± 1 SD.