MIGRAINE CLASSIFICATION

Balu S Unny
Rajagiri College of Social Sciences, Kalamassery

Abstract - Migraine is a medical condition responsible for severe recurring headaches that can throb or pulse. Symptoms responsible for the condition include nausea, neck pain, dizziness, nasal congestion, vomiting, and extreme sensitivity to light and sounds. A Migraine episode can usually last for many days and can affect the daily life of a person which includes the ability to work or study or carry out many other tasks. A Term called Aura can occur, including physical or sensory symptoms such as flashing lights. The aim of this project is to classify which type of migraine occurs for each variable. Data's include various symptoms or auras from many patients, which is used to classify the type of migraine affected for each medical patient. Researchers from 2018 found out that more than 15% of adults in the US have experienced migraine episodes or a headache with high severity within the last three months. Migraine Frequency can be once a year, once a week or any amount within that period. The Most common occurrence of migraine is between two to four times per month. It is important to understand what factors are responsible for a particular type of migraine The project aims to analyze the dataset using visualization and build multiple classification models using Orange and to select the model with the best accuracy.

Keywords: **Migraine, Classification, Machine Learning, Algorithms, Preprocessing, Exploration**

I. INTRODUCTION

Migraine is a common neurological disorder that affects millions of people worldwide, causing significant disability and reduced quality of life. One of the challenges in managing migraines is the classification of different types of migraines, which can vary in their symptoms, triggers, and response to treatment. Traditionally, migraine classification has been based on clinical criteria, but recent advances in machine learning algorithms have shown promising results in improving the accuracy of classification. The condition can significantly impact the quality of life of those affected, causing severe pain, nausea, and sensitivity to light and sound. The classification of migraines has been a long-standing challenge in the medical community, as there are various subtypes with differing symptomology and treatment options. In this Research paper we will explore the dataset of Migraine to train and test several machine-learning models, including Logistic Regression, KNN, Naive Bayes,SVM, Gradient Boosting and evaluate the performance of each model using metrics such as accuracy, precision, recall, and F1-score.

II. LITERATURE REVIEW

[1] In a study published in the journal Cephalalgia, researchers used a machine learning algorithm called random forest to classify migraines based on their clinical features. The algorithm was trained on data from 3,000 patients and achieved a classification accuracy of 92.8%. The researchers concluded that machine learning algorithms could improve the accuracy of migraine classification and assist in personalized treatment selection.

[2] Another study published in the journal Neurology Genetics used a machine learning algorithm to identify genetic risk factors for migraine. The algorithm analyzed genomic data from 59,674 individuals and identified several genetic variants associated with migraine. The researchers concluded that machine learning algorithms could be used to identify new genetic targets for migraine therapy.

[3] A review article published in the journal Headache summarized several studies that have used machine learning algorithms to classify migraines. The authors concluded that machine learning algorithms have shown promise in improving the accuracy of migraine classification and could be used to develop personalized treatment strategies.

[4] In a study published in the journal Pain Medicine, researchers used a machine learning algorithm to predict the response to migraine treatment based on patients' clinical features. The algorithm was trained

on data from 221 patients and achieved a classification accuracy of 84.4%. The researchers concluded that machine learning algorithms can be used to provide treatments for patients.

[5] A review article published in the journal Frontiers in Neurology discussed the potential of machine learning algorithms in predicting migraine attacks. The authors concluded that machine learning algorithms could be used to develop predictive models for migraine attacks based on patients' clinical features, environmental factors, and other factors. These predictive models could assist in developing personalized preventive strategies for migraine.

III. IMPLEMENTATION

Orange is an open source component based tool software based on python which is used for data visualization,data mining and machine learning. It is a platform developed for experiment-based selection, predictive modeling and recommendation system.

A) Data Description

The Migraine Classification dataset was obtained from kaggle. The Following attributes are used to classify the type of migraine based on the symptoms. The Dataset consists of 400 instances and 24 features. Out of 24 features there exists 16 categorical and 8 numeric variables.

The attributes in the dataset are:-

- Age
- Duration
- Frequency
- Location
- Character
- Intensity
- Nausea
- Vomit
- Phonophobia
- Photophobia
- Visual
- Sensory
- Dysphasia
- Dysarthria
- Vertigo
- Tinnitus
- Hypoacusis
- Diplopia
- Visual defect
- Ataxia
- Conscience
- Paresthesia
- DPF
- Type

No	Attribute	Description	Type
1	Age	Patient’s Age	N
2	Duration	Symptom Duration in Days	N
3	Frequency	Episodes per month	N
4	Location	Pain Location	N
5	Character	Pain Type	N
6	Intensity	Intensity of the pain	N
7	Nausea	Nauseous feeling	C
8	Vomit	Stomach Uneasiness	C
9	Phonophobia	Noise Sensitivity	C
10	Photophobia	Light Sensitivity	C
11	Visual	Reversible visual symptoms	N
12	Sensory	Reversible sensory symptoms	N
13	Dysphasia	Lack of Speech Coordination	C
14	Dysarthia	Disarticulated sounds and words	C
15	Vertigo	Dizziness	C
16	Tinnitus	Ringing in the ears	C
17	Hypoacusis	Hearing loss	C
18	Diplopia	Double Vision	C

19	Defect	Frontal eye and nasal field defect	C
20	Ataxia	Lack of Muscle Control	C
21	Conscience	Jeopardized Conscience	C
22	Paresthesia	Simultaneous bilateral paresthesia	C
23	DPF	Family Background	C
24	Type	Diagnosis of Migraine	C

B) Data Preprocessing

The Dataset 'Migraine Classification' did not have any missing values which resulted in less preprocessing steps. In the following dataset the feature 'age' was selected and discretized to an interval of convenience.This was done to understand the age group which was prone to the particular migraine type.

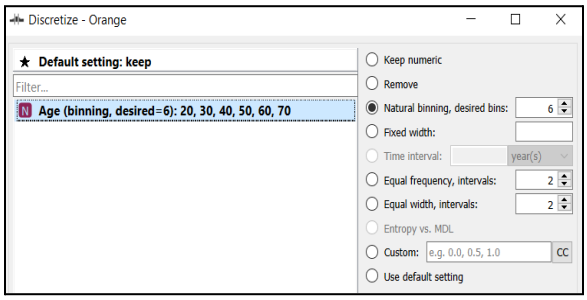


Fig: Age discretization

	Age
1	30 - 40
2	50 - 60
3	50 - 60
4	40 - 50
5	50 - 60
6	40 - 50

Fig: Discretized data table of age

C) Data Exploration

Data exploration is an essential step in data analysis that involves examining and understanding the structure and characteristics of the data set. It is a critical process that helps in identifying patterns, trends, and relationships within the data, which can be used to develop insights and inform decision-making.

In data exploration, researchers typically start by examining basic statistics such as the mean, median, mode, and range of the variables in the data set.

Next, researchers can use visualization techniques to explore the data graphically. This can include creating scatter plots, histograms, box plots, and other visualizations that help to identify patterns and relationships between variables. These visualizations can also help identify any data quality issues, such as missing or incomplete data, that need to be addressed.

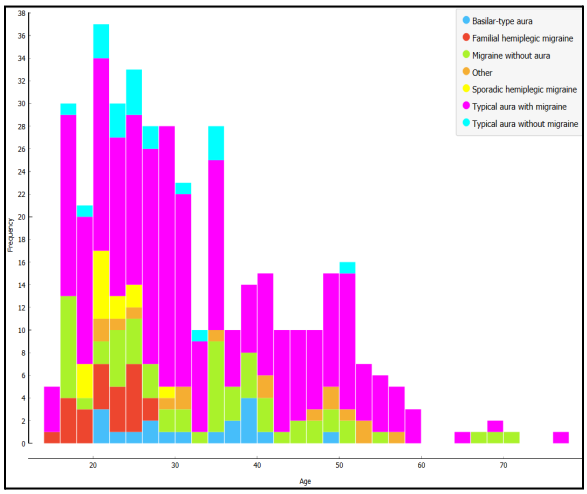


Fig: Distribution of Age

The Above fig depicts the distribution of patient's age w.r.t the type of migraine occurred. It is evident that 'typical aura with migraine' is the most frequently occurring migraine type out of the total records with ages between 20 to 40.

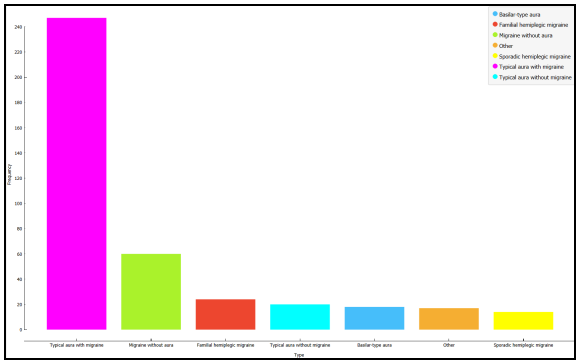


Fig: Distribution of Migraine Types

It can be conluded from the above figure that the 'typical aura with migraine' occurred the most which is 247 out of the 400 records followed by other migraine types with a frequency of less than 60.

D) Classification Technique

The Classification technique is a technique for organizing data points into a variety of categories. The Classification technique is done with the help of Machine leaning tools.The Main purpose of classification is to dig valuable information based on the data. This model predicts class labels and tests the constructed model based on test data, hence estimating the classification rules. The Project uses the classification techniques such as Logistic Regression, K-Nearest Neighbour, Naive Bayes Classification, Support Vector Machine, Gradient Boosting. The Target variable here is 'Type' which has contains 7 of migraine to be classified that includes 'Typical aura with migraine','Migraine without aura','Typical aura without migraine','Familial hemiplegic migraine','Sporadic hemiplegic migraine','Basilar-type migraine' and other. These types are encoded into values in order to give as input to the models for ease of use with values from 0 to 7.

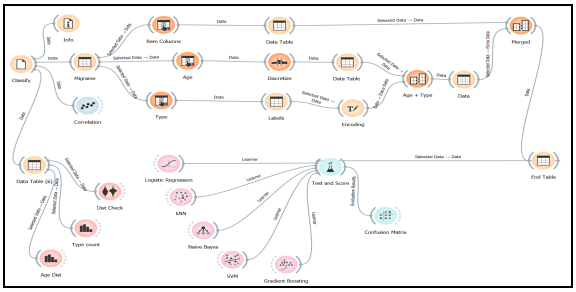
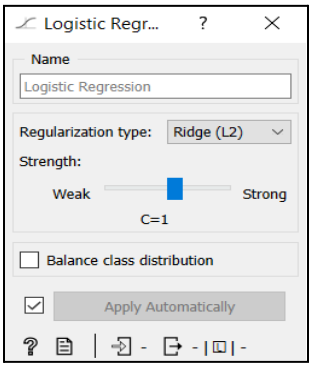


Fig: Model Building

i) Logistic Regression

Logistic regression one of the most popular machine learning algorithms is a supervised learning technique which is used to predict the output of a categorical dependent variable.



The model evaluation value of Logistical Regression are shown below:

Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.975	0.900	0.891	0.892	0.900

Fig: Logistical Regression Evaluation Parameter

ii) K-Nearest Neighbour (KNN)

The K-Nearest Neighbours algorithm also known as KNN is a supervised leaning technique that uses proximity inorder to make classifications of an individual data points. It is mostly used as afor classification even though it is also used for regression problem.

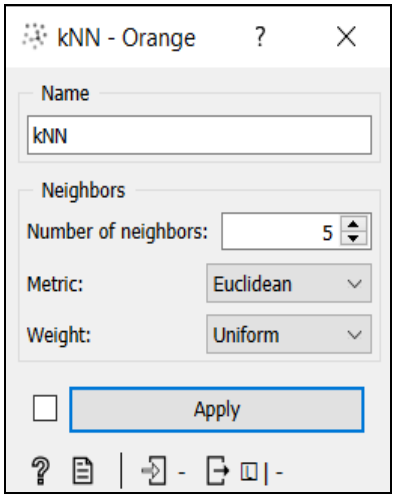


Fig: KNN Model Parameters

The model evaluation value of Logistical Regression are shown below:

Model	AUC	CA	F1	Precision	Recall
kNN	0.899	0.810	0.762	0.788	0.810

Fig: KNN Model Evaluation Parameter

iii) Naive Bayes Classifier

Naive Bayes is a type of probabilistic machine learning model which is used for the classification task with an assumption among predictors. The Naive Bayes assumes that the presence of a feature is unrelated to another feature. It is easy to build and one of the useful algorithms for large data sets.

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

where:

P(A|B) = Conditional Probability of A given B
P(B|A) = Conditional Probability of A given B
P(A) = Probability of event A
P(B) = Probability of event A

Fig: Naive Bayes Formula

The model evaluation value of the Naive Bayes Algorithm is shown below:

Model	AUC	CA	F1	Precision	Recall
Naive Bayes	0.985	0.908	0.912	0.920	0.908

Fig: Naive Bayes Model Evaluation Parameter

iv) Support Vector Machine (SVM)

Support Vector Machine also known as SVM is used for both classification or regression problems. It is one of the most popular algorithms used for text classification. In Support Vector Machine, each data item is plotted as a point in n-dimensional space.Then the classification is done by finding the best hyperplane, which differentiates the two classes very well.

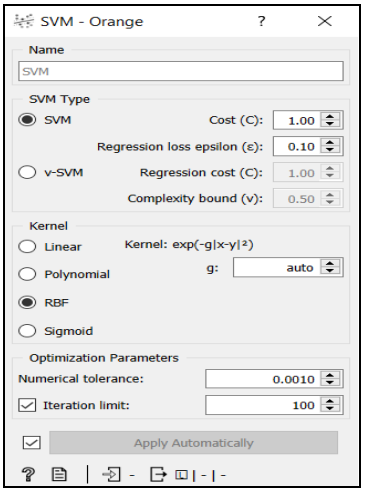


Fig: SVM Model Parameters

The model evaluation value of the Naive Bayes Algorithm is shown below:

Model	AUC	CA	F1	Precision	Recall
SVM	0.960	0.871	0.859	0.860	0.871

Fig: SVM Evaluation Parameter

v) Gradient Boosting

Gradient Boosting is a group of Machine Learning algorithms that combines all the weak models together to create a strong predictive model.Decision trees are the ones that are usually used when doing gradient boosting.It is highly known for classifying complex datasets.Each tree predicts a label and the end prediction is done with an equation.

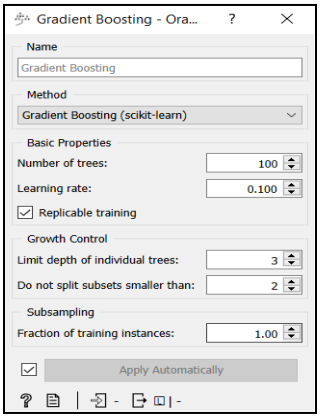


Fig: Gradient Boosting Model Parameters

The model evaluation value of the Gradient Boosting is shown below:

Model	AUC	CA	F1	Precision	Recall
Gradient Boosting	0.978	0.893	0.887	0.886	0.893

Fig: Gradient Boosting Evaluation Parameter

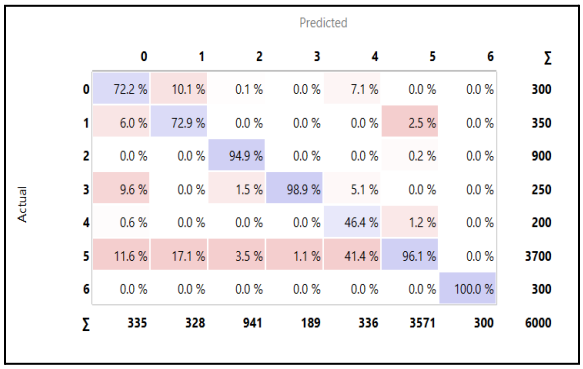
IV) RESULT AND DISCUSSION

The Dataset 'Migraine Classification' consists of 400 rows and 24 columns where the selected target variable is 'Type'. The Data is split into 70:30 ratio for train and test.

The Models applied to the dataset are Logistic Regression, KNN, Naive Bayes,SVM, Gradient Boosting.The Performance metrics for the each models are given below:

Model	AUC	CA	F1	Precision	Recall
kNN	0.899	0.810	0.762	0.788	0.810
SVM	0.960	0.871	0.859	0.860	0.871
Naive Bayes	0.985	0.908	0.912	0.920	0.908
Logistic Regression	0.975	0.900	0.891	0.892	0.900
Gradient Boosting	0.978	0.893	0.887	0.886	0.893

Among the classification models, Naive Bayesian classifier has the highest accuracy of 90.8%. The Confusion Matrix of the classifier is given below.



V. CONCLUSION

In this paper a methodology was proposed for predicting the type of migraine based on various symptoms from medical records of 400 patients. These symptoms are essential in order to predict the type of migraine. The Dataset cosnisting of patient's data was used to classify using several machine learning models which are Logistic Regression, KNN, Naive Bayes,SVM, Gradient Boosting etc. These conclusions can be used in the future to be aware of the condition when these symptoms or aura shows up right before the occurence of migraine.The Analysis concludes that the most symptoms are associated with the 'typical aura with migraine' which is the most frequently occured. Among the applied classification techniques Naive Bayes Classifier gave the best accuracy.

REFERENCES

[1] Ahmed, A. E., & Khan, F. A. (2020). Migraine classification using machine learning algorithms: A systematic review.

[2] Ardiç, Ü. A., & Alkan, A. (2019). Classification of migraine patients using machine learning algorithms. Medical & Biological Engineering & Computing

[3] Capi, M., Corbelli, I., Gozzi, N., Favoni, V., & Pierangeli, G. (2020). Machine learning techniques in migraine classification: A review. Journal of Headache and Pain

[4] Díaz-de-Mera, Y., Roldán-Blasco, C., Alberola-López, C., López-de-Ullibarri-Gómez, A., González-Sánchez, M., & García-Martí, G. (2019). Machine learning classification of migraine without aura: An application on neuropsychological tests.

[5] Goadsby, P. J. (2020). The importance of migraine classification. Cephalalgia

[6] Pogacnik, T., & Pirc, K. (2020). Machine learning in migraine diagnosis and treatment selection.

[7] Yuan, H., Xu, J., Chu, Y., & Chen, N. (2020). Identifying migraine patients using support vector machine and random forest based on magnetic resonance imaging.