iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** I **Month of publication:** January 2023

DOI: https://doi.org/10.22214/ijraset.2023.48840

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue I Jan 2023- Available at www.ijraset.com

Quality Prediction Model for Drug Classification using Machine Learning Algorithm

Ayush Kumar
Symbiosis Institute of Operations Management

Abstract: ***Thousands of approved drugs are used to treat people who suffer from various medical problems. Therefore, adverse effects and other protection uncertainty are important to be recognized that can help in patient control. The objective of this paper is to build and compare different machine learning models for classifying of drugs. The dataset comprises of the data about a set of patients and their response to one of the five medications. This paper will investigate which algorithm generates the most accurate prediction for Drug classification.***
Keywords: Drugs Classification, KNN, Support Vector Machine, Random Forest, Decision Tree.

I. INTRODUCTION

A great many infections undermine the prosperity of the people and consistently new ones are being added to the current number of sicknesses. There are infections which have no fix and has tormented the populace for quite a long time. Consequently, quickly and precisely discovering the medications than can successfully treat or ease the sicknesses is profoundly basic. There are many advances that are needed from improvement of medications to definite stockpile of medications. These incorporate preclinical and clinical preliminaries of these medications. The overall accomplishment speed of prescription disclosure and preclinical assessments, which are fundamental for the lab improvement stage, is generally 0.05–0.1%, and under 1% of the candidate compounds are presumably going to have the ordinary effect and proceed to the clinical starter stage. Thusly, examination of medication to target cooperation is a vital necessity during the time spent disclosure of medications and this can work on the pace of achievement of revelation of new medication. There isn't just a need to use significant assets to look and test the competitor intensifies individually in the length of the advancement period of the medication to certify that they meet the assumptions, yet it is likewise to exhibit the meaning of medication to target association forecast in the entire interaction of the medication improvement. The downside of testing bio therapeutically to discover the compound is that it doesn't uphold quick finding and taking care of issues which will be negative for the treatment of arising and profoundly irresistible illnesses. Thusly, in the forecast of medication to target communications, AI methods have been presented. The calculations of AI do the information investigation and assemble a model utilizing the datasets to create prescient models and in this manner, has become a fundamental strategy for natural exploration.

II. LITERATURE SURVEY

In medical applications, such as detecting the type of malignant cells, machine learning is commonly used. One of the most commonly found cancers i.e. the breast cancer, accounts for a large number of fatalities every year. Cells that are cancerous are classified as malignant (M) benign (B) or cancerous cells (M). Decision Tree, Support Vector Machine (SVM), k Nearest Neighbours and Naive Bayes (NB) are some of the methods used to classify and forecast breast cancer. The given dataset is trained with these algorithms and then the prediction accuracy of each technique is compared. (Bharat, Pooja, & Reddy, 2018). One of the most researched topics in medicine is cancer diagnosis. Many researchers have focused their efforts in order to improve the performance and generate good results in the diagnosis of cancer. In the research of cancer, the diagnosis of breast cancer is a major issue. System learning is a discipline in artificial intelligence that supports the evolution of machine as per the process. Machine learning is utilised in bioinformatics, especially in the diagnosis of breast cancers. K-nearest neighbours, a method in supervised learning, is a method which is very common. It's fascinating to think of using the K-NN in medical diagnosis. The parameter "k," which represents the count of neighbours that are nearest, have a big impact on the quality of the findings. (AhmedMedjahed, Ait Saadi, & Benyettou, 2013). When it comes to detecting lung cancer nodules, machine learning algorithms are contrasted. To detect the anomaly, machine learning approaches like support vector machine, decision trees and KNN are used. All of the techniques are compared both without and with pre-processing. (Günaydin, Günay, & Şengel, 2019). The SVM-KNN classifier's performance in detecting breast cancer using a tumour dataset is investigated.

©IJRASET: All Rights are Reserved

The goal is to determine if a tumour is benign or malignant based on cell descriptions obtained from microscopic analysis. The SVM-KNN classifier's classification performance is examined and compared to that of the support vector machine. The SVM-KNN model produced a great performance with 98.06 percent classification accuracy on the testing subset, according to the results (Li & Sun, 2010). KNN, NN, and SVM classification algorithms were used to suggest a method for automating the segmentation, feature extraction, and classifying of red and white blood cells for Leukemia (Sachin & Kumar, 2017). On the dataset gathered from the UCI repository, support vector machine and KNN are used to diagnose breast cancer. The efficiency and the accuracy of algorithms are also tested and compared (Saturi, Scholar, Sai Phani, & Chand, 2021). Multiple classifiers were employed and their relative accuracy parameters were compared to to build and design a structural approach for the brain tumour prediction at an earlier stage using different machine learning algorithms. (Gupta, Sharma, Saxena, & Arora, 2021). KNN classification is performed by using the interaction labelled data of the user and which spatial label dependency models were probed for the problem of the segmentation of brain tumour (Havaei, Jodoin, & Larochelle, 2014). To find the classification model, three traditional algorithms are used: logistic regression, KNN and SVM. The methods were then employed to forecast categorization model. After developing the model, a more efficient classification model is discovered based on the metrics. Whether the data is benign or malignant, the suggested method provides a more satisfactory outcome for tumour classification (Kalaiyarasi, Dhanasekar, Sakthiya Ram, & Vaishnavi, 2020). On the datasets of cancers involving Leukemia , the Colon , and the Lymphoma , structure adaptive self–organizing map, KNN and SVM were employed to classify for prediction and diagnosing cancer (Cho & Won, 2003). Breast cancer diagnosis and prognostic risk are being investigated. SVM, KNN, and PNN are three well-known classifiers used to assess recrudescence and metastasis (Osareh & Shadgar, 2010). Machine learning techniques are used to detect tumours in brain MRI. Pre-processing processes are conducted to MRI images of brain, features regarding texture are retrieved by using Gray Level Co-occurrence Matrix, and by using a machine learning method the classification is performed (Sharma, Kaur, & Gujral, 2014). A review of numerous research articles is conducted for the accuracy comparison of various algorithms regarding machine learning which are based on datasets and attributes provided of the disease cancer. Random Forest, Decision Trees, SVM, KNN, Fuzzy Neural network, Artificial neural network and others are used in the prediction modelling. Among the various machine learning approaches, SVM provides the most accurate results for predicting the cancer for the given dataset (Kumar, Sushil, & Tiwari, 2019). The use of data mining techniques is being investigated in attempt of the improvement in the accuracy of survival prediction of cancer. The Python programming language is used to build three data mining methods on the data sets: Random Forest, Decision Tree, and K-Nearest Neighbors, and the computed accuracy prediction data rates are proportionate to the existing approaches. The study's findings have been proven to be very good for breast cancer projections. The suggested technology can quickly determine the stage of the disease is currently in and predict whether it will be malignant or not for the patient (Ghosh & Hasan, 2020). An automatic method for classifying brain tumours as benign vs. malignant and high grade vs. low grade glioma is presented. The texture features from photos are extracted using the GLCM approach and is stored as feature vector in this method. Supervised KNN and SVM algorithms were used for classification of the extracted features. The suggested approach is tested on 251 images from the clinical database (166 benign and 85 malignant) and 80 images from the brats 2012 training database (30 high grade glioma and 50 low grade glioma). For the clinical database, the suggested system's accuracy registers 86 percent and 96 percent for KNN and SVM respectively, and 72.5 percent and 85 percent for KNN and SVM for the database of Brats (Wasule & Sonar, 2017). Chronic nephritis also known as chronic kidney disease, is a condition where the kidneys are affected resulting in limitations in your ability to remain healthy. Increased blood levels, anaemia, nerve damage and weak bones are all possible complications. The accuracy, precision, and execution time of the KNN, Random Forest and Naive Bayes classifiers for predicting chronic kidney disease were evaluated (Devika, Avilala, & Subramaniyaswamy, 2019). To predict the breast cancer severity, four machine learning approaches are tested and implemented on mammography patients dataset: KNN, ANN, SVM and Decision Trees. This study makes a contribution by assessing all the given models and identifying the best system which is based on a various evaluation metrics like sensitivity, accuracy, specificity, etc. (Laghmati, Tmiri, & Cherradi, 2019). Machine learning has made an appearance in the medical field, with the goal of offering tools and evaluating data connected to diseases. As a result, algorithms of machine learning are critical for attaining early detection of diseases. This research reviewed numerous machine learning techniques for illness detection. Standard datasets have been utilised in diseases such as liver, chronic renal disease, breast cancer, heart disease, brain tumours, and many more (Ibrahim & Abdulazeez, 2021). This article discusses Machine Learning and how it can be useful for detecting and investigating cancer tumours. Simple procedures were utilised to build a powerful machine learning programme that can determine if a tumour is malignant or benign. This strategy can be used by nearly anyone thanks to Python and its open source libraries. To compare the findings and select the more accurate algorithm, the KNN classifier and Logistic Regression are employed to create two models (Agarwal & Saxena, 2018).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue I Jan 2023- Available at www.ijraset.com

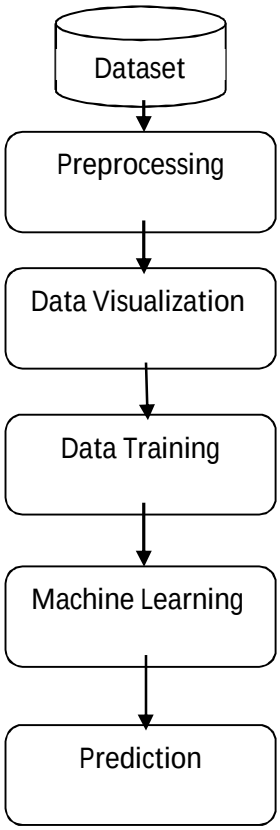
For effective breast cancer detection, the paper presents a hybrid model combining numerous machine learning methods, including ANN, SVM, KNN and decision tree. The datasets utilised for detection of breast cancer and diagnosis are discussed in this paper (Tahmooresi, Afshar, Bashari Rad, Nowshath, & Bamiah, 2018). In the breast cancer dataset of UCI Wisconsin, the accuracy of classification of three Machine Learning algorithms – kNN, ANN, and NB – is compared. The goal of this comparative study was to determine which machine learning method had highest accuracy for diagnosing breast cancer (Neagu, Guo, Trundle, & Cronin, 2007).

The influence of feature selection approaches, employs a approach of a filter, on the error and accuracy of supervised cancer classifications are investigated in this research.

A comparison of multiple selection approaches like T- Statistics, SNR etc. is conducted on a dataset of malignancies including leukaemia, colon and prostate cancers.

Results of the classification using SVM and KNN classifiers demonstrate that the SNR's approach combined with the SVM classifier provides the most accurate predictions.(Bouazza, Hamdi, Zeroual, & Auhmani, 2015).

Flow of work



The dataset Drug Classification has been taken from Kaggle.

This data is collated from a drug company which comprises of dataset of the drugs (labelled) and the parameters effecting it.

Therefore, by analyzing on the basis of disease and the type of patient, the recommendation of the drug is done.

III. DATA PREPROCESSING

The process of preparing of the raw form of data, to train and test the data suitable for machine learning is called data preprocessing. It is the most important step when the machine learning model is being created. In any dataset or any real world data, there are a missing values, contains noise and sometimes is in a format that is unusable for the machine learning processes. Data preprocessing consists of tasks that cleans the dataset and makes it suitable for the machine learning process and also drastically increases the accuracy making it more efficient for machine learning process.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

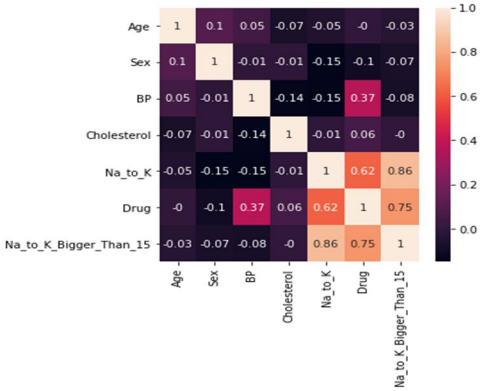
A. One Hot Encoding

There are some machine learning techniques that can work only when categorical data has been provided. In a decision tree, no data transformation is required and it can be learned directly from the categorical data. There are algorithms that are unable to operate directly on the label data. Numeric input and output variables are required for these types of algorithms. In general, this is usually a constraint of an efficient machine learning technique and not a limitation on the algorithms. This implies that it is necessary to convert categorical data into a numerical form. If the output variable is categorical, it is good to convert predictions suggested by the model back to a categorical form so that is can be presented or useful in the application.

In the case where categorical variables are used, there is no ordinal relationship and therefore the method of integer encoding is insufficient. There will be a poor performance if we use integer encoding as it will permit the model to go through a natural order between all the categories. Addition of a new binary variable is done for each integer value and when integer gets encoded, the variable is removed.

B. Correlation Check

The correlation check is an essential factor as it finds the features that are highly correlated among themselves that results in high redundancy and might result in poor predictions. The correlation can be checked and estimated by using 'heatmap'. The heatmap visualization is used as it provides clarity among the correlation features.

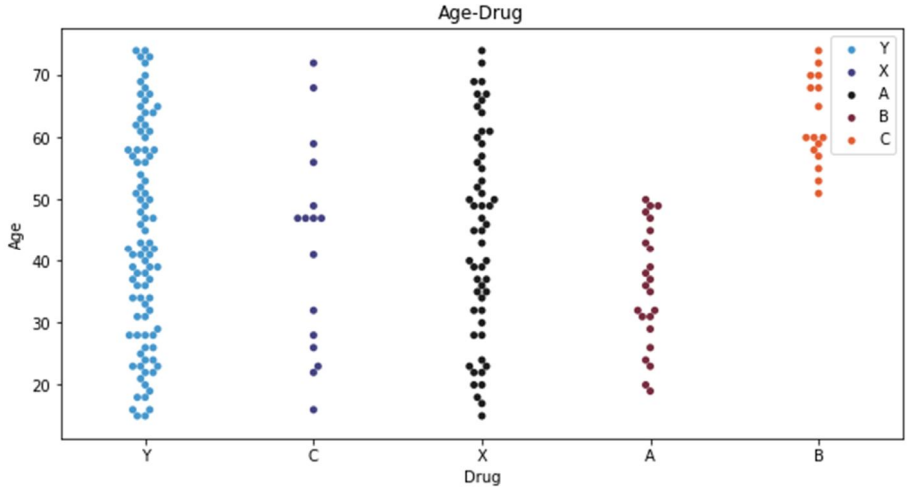


Visualization of correlation using Heatmap(fig)

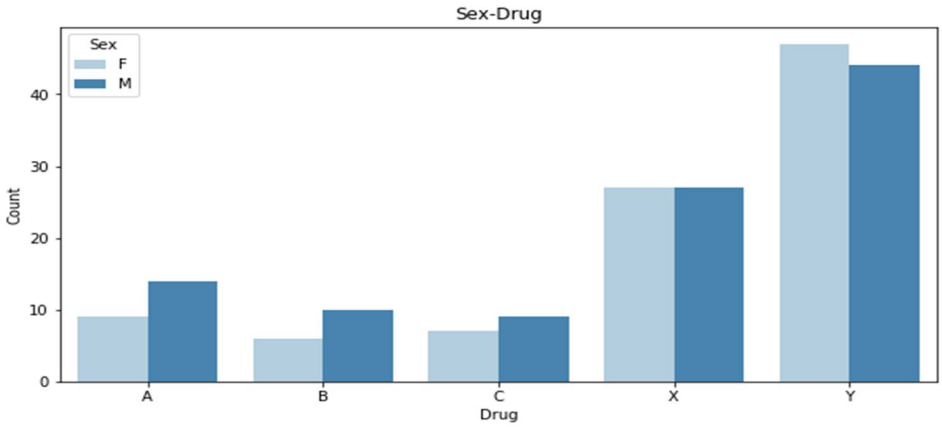
The statistical metric that is used for measuring to what degree the different variables are interdependent is known as correlation. Here, high cholesterol is directly linked to having high blood pressure. Therefore, cholesterol is known to be correlated to blood pressure since higher the cholesterol higher the blood pressure.

IV. DATA VISUALIZATION

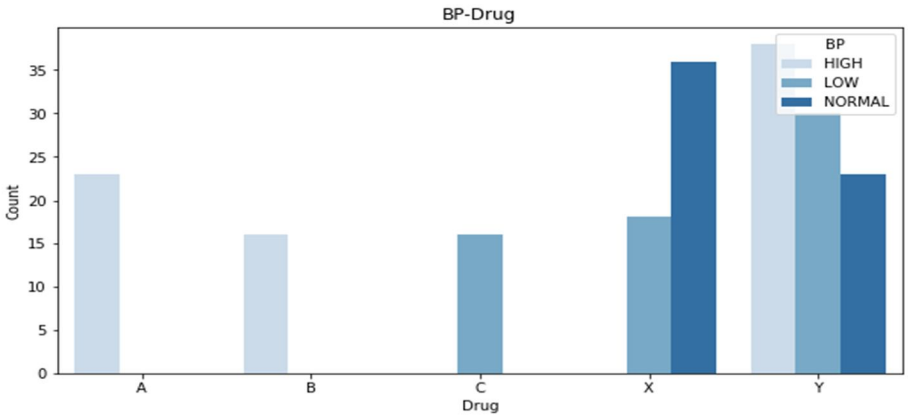
The above graph shows the age distribution with respect to drug



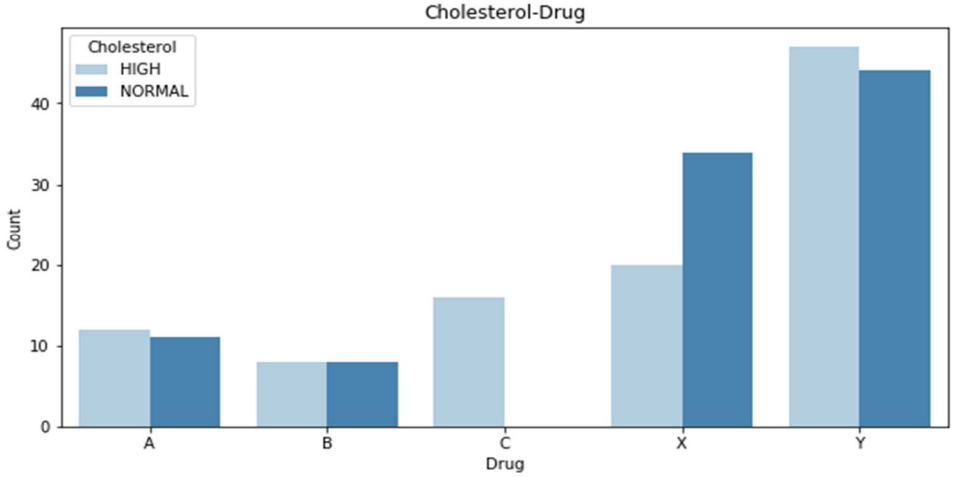
From the above visualization, the distribution of people is done according to the drug they are recommended to consume among the five drugs. For example, it can be seen that the Drug B is to be taken only by person who is older than 51 years old and Drug A is to be taken by person who is younger than 50 years old.



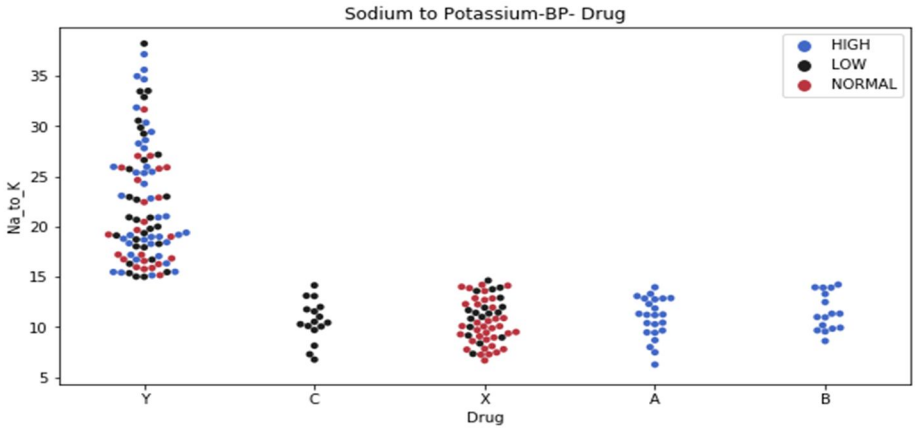
In the above visualization, the distribution of males and females to drug has been done. It can be seen that the count of females getting Drug A, Drug B, Drug C is less in number than the males. Females get Drug Y more than the males whereas the number of males and females receiving Drug X is equal. Sex feature does not seem to be an essential feature for classification.



In this visualization, blood pressure to drugs is compared. Here it can be observed that Drug A and Drug B is only taken by the people who have high blood pressure. Drug C is taken by people who have low blood pressure and Drug A is taken by the people who have high blood pressure. Here, it can be seen that blood pressure is an important feature for classification.



In this visualization, cholesterol is compared to drugs. It can be observed that Drug C is taken only by the people who have high cholesterol. This is an important feature to classify Drug C.



In the above visualization, a swarm graph is shown with features of ratio: potassium to sodium, drug to blood pressure. People who have blood pressure at elevated levels and less than 15 as potassium to sodium ratio, take Drug A and Drug B only. The people who have low blood pressure and less than 15 as potassium to sodium ratio, they take Drug C only.

V. TYPES OF REGRESSION MODELS

A. K nearest Neighbor Classifier

The K-nearest neighbor classifier is a supervised machine learning algorithm which is useful for classification and regression prediction of problems. In this case of drug classification, it is used for classification purpose of prediction of drugs. There are two properties that describe the KNN algorithm: -

1) Lazy Learning: Utilizes all the values of the dataset for training while classifying and lacks specialized train phase

2) Non- parametric Learning: There are no assumptions about the underlying data.

B. Random Forest (RF) Classifier

Random Forest is a type of supervised learning algorithm that finds usage in classification and regression, though it mainly finds usage in the classification type problems. Based on the data samples, the random forest algorithm generates decision tree and then it uses each of these samples to generate predictions. Then it selects the best solution among the predicted vales which is done by the method of voting.

C. Support Vector Machine (SVM) Classifier

Support Vector Machine (SVM) is a type of supervised machine learning algorithm that finds usage in both the classification and regression type problems, though it is usually useful in classification type problems. In support vector machine algorithm, plotting of each item of the dataset is done as a point in an n-dimensional space where the value of a particular coordinate is equivalent to the value of each feature. Here n is the count of features that exists. Then, classification is performed by detecting the hyper-plane that differentiates the two classes.

D. Decision Tree classifier

The decision trees are made by dividing the training set of the dataset into distinct nodes, where each node contains most of all the information regarding the category. In this case, we are prescribing a drug to a patient, but the situation of the patient's biology is going to determine which drug to take i.e. Drug A or Drug B. This can be determined by many features. One such feature is the age of the patient. The patient can be a young, senior or a middle aged person. If the patient is a middle aged person, then Drug B will definitely be chosen. In the case where the candidate is a young person or a senior person, more details will be required to determine the drug prescription to the respective individual. The other features generally are variables like gender or the level of cholesterol. For example, if the candidate is a female patient, then Drug A will be recommended but if the candidate is a male patient, then Drug B will be recommended.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue I Jan 2023- Available at www.ijraset.com

E. Hyper Parameters Tuning using Grid Search

Tuning the hyper parameters is important as they control the overall functioning of a machine learning model. The final goal is to achieve an optimal combination of these parameters that will minimize the predefined loss function to give improved results.

Here, I have used grid search method to tune the hyper parameters. In this technique, a model is made for each possible combination of all the hyper parameters values that is given. Evaluation is done of each model and then a final model is selected which has the architecture that produces best results.

There are some parameters that are not included in the process of training and fitting the dataset. These parameters are known as hyper parameters. Scikit-learn uses cross-validation to automatically iterate over these hyper parameters. This is also known as the Grid Search method. Grid Search uses the objects that are required to train and different values of the hyper parameters. Then mean square error or R-squared value is calculated for various hyper parameter values, allowing the user to choose the values that give best accuracy for training and fitting of the dataset.

VI. TYPES OF ERRORS

A. Mean Square Error (MSE)

The mean square error is the most common and the simplest loss function in machine learning. The calculation of mean square error is done in three steps. They are:-

1) Taking the difference of the model's predictions and the ground truth
2) Squaring the difference of the above step
3) Averaging the value across the whole dataset

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Where N is the number of samples used in testing. The symbol sigma is denoting the difference between the actual and predicted data vales which is being taken on every j value which ranges from 1 to n.

Mean square error is a great method to ensure that the trained model is not having any outlier predictions with vast errors, as it puts more weights on the errors by squaring the function part. The disadvantage of this method is if the model has a single very bad prediction, squaring that part of the function will magnify the error. In practical scenarios, this is usually ignored as the model which is being made is expected to perform good on the majority and not on the outliers. Since the errors are always squared, mean squared error cannot be negative. Below is the equation of mean squared error:

B. Mean Absolute Error (MAE)

The mean absolute error has a very slight difference when compared to the mean squared error. The calculation of mean absolute error is done in three steps. They are:-

1) taking the difference of the model's predictions and the ground truth
2) applying the absolute value to the result of the above step
3) averaging the value across the whole dataset

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

Where n represents the number of observations of the dataset. The symbol sigma is denoting the absolute difference between the actual and predicted data vales which is being taken on every j value which ranges from 1 to n.

The mean absolute error covers up for the mean squared error disadvantage. As the absolute value is taken, each error is weighed on the same linear scale. This is different than mean squared error where outliers were given too much weight. The disadvantage will be if outliers are important in the model, then this method is not very effective. If there will be large errors from the outliers, the mean absolute error will weigh the same errors as low errors which will result in poor prediction of the model.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. R-squared Error

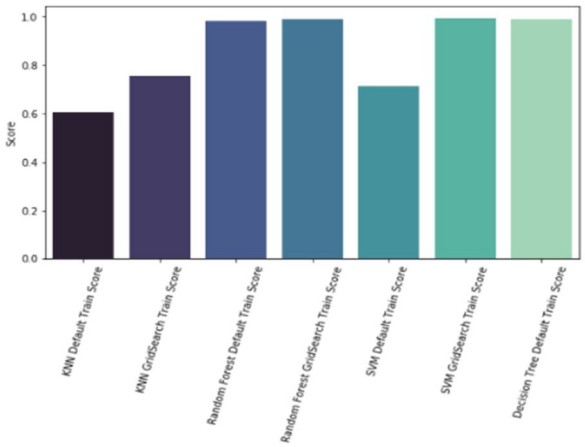
R-squared error is also called as coefficient of determination. The indication about how good a model can fit a dataset is determined by the metric of coefficient of determination. It is basically indicating how close the predicted values are plotted to the actual values of the given dataset. The value of the R- squared lies in the range of 0 to 1. If the value is near zero, it indicates that the model does not fit the given dataset. The more the value is near to 1, the more the model will fit perfectly with the dataset that is provided.

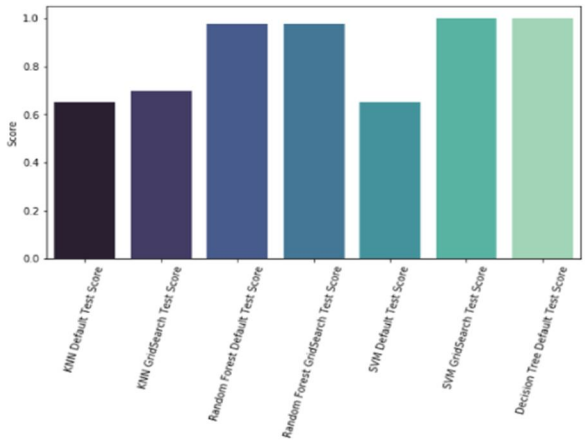
VII. RESULTS

After the raw data is preprocessed and transformed into useful data, the dataset is split into two parts. The first part of split dataset is the training set which comprises of 80% of the dataset and the second part is the testing set which comprises of 20% of the dataset. Comparative analysis of various algorithms with respect to training and testing score is given below:

	Score
KNN Default Train Score	0.606250
KNN GridSearch Train Score	0.756250
Random Forest Default Train Score	0.981250
Random Forest GridSearch Train Score	0.987500
SVM Default Train Score	0.712500
SVM GridSearch Train Score	0.993750
Decision Tree Default Train Score	0.987500

	Score
KNN Default Test Score	0.650000
KNN GridSearch Test Score	0.700000
Random Forest Default Test Score	0.975000
Random Forest GridSearch Test Score	0.975000
SVM Default Test Score	0.650000
SVM GridSearch Test Score	1.000000
Decision Tree Default Test Score	1.000000





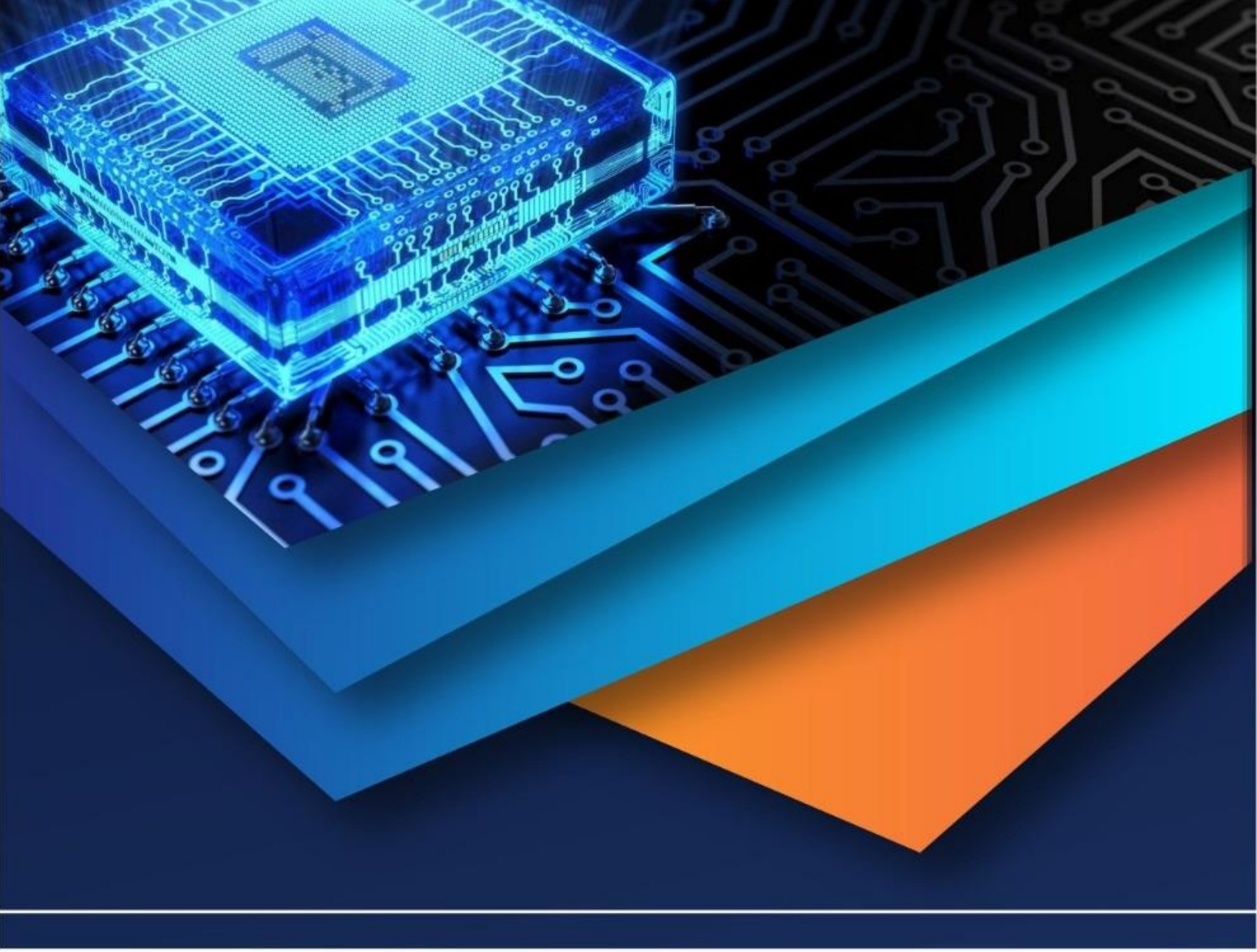
The performance of the model and the precision of the model training will be determined by the accuracy of the training set. The accuracy of the training dataset will determine how well the model performs during the real prediction. Random forest classifier and SVM classifier have more accuracy and results in good score after hyper parameter tuning. KNN default train score has the worst accuracy among all the models tested for training. Despite being very inaccurate in default test score, SVM classifier after hyper parameter tuning gives a cent percent accuracy in test scores. The Decision tree classifier gives a perfect test score and therefore, hyper parameter tuning has not been done on it. Another reason of not hyper tuning the decision tree classifier is due to the reason of overfitting. In the cases where the training set of data is very well trained, the classifier will pick up outliers and disturbances of the dataset and give wrong prediction as the model is overly trained. It is therefore important to have a classifier that does not give either overfitting or under fitting results. Hence, it can be concluded that the Decision Tree classifier is the best classifier among the given classifiers to recommend drugs.

VIII. LIMITATIONS

Machine learning algorithms have played an important role of drug discovery and classification. These techniques increase the efficiency and explore hundreds of combinations which would not have been possible without this technology. Although, machine learning is effective, there are some limitations. The data about biological effect of specific protein is limited which results in less extrapolated data. The effect of a certain drug on the drug used is not observed and therefore there is a lack of drug to drug interaction and effect on the health of the patient.

REFERENCES

[1] Agarwal, A., & Saxena, A. (2018). Malignant Tumor Detection Using Machine Learning through Scikit-learn, 119(15), 2863–2874. Retrieved from http://www.acadpubl.eu/hub/
[2] AhmedMedjahed, S., Ait Saadi, T., & Benyettou, A. (2013). Breast Cancer Diagnosis by using k-Nearest Neighbor with Different Distances and Classification Rules. International Journal of Computer Applications, 62(1), 1–5. doi:10.5120/10041-4635
[3] Bharat, A., Pooja, N., & Reddy, R. A. (2018). Using Machine Learning algorithms for breast cancer risk prediction and diagnosis. 2018 IEEE 3rd International Conference on Circuits, Control, Communication and Computing, I4C 2018, (x), 1–4. doi:10.1109/CIMCA.2018.8739696
[4] Bouazza, S. H., Hamdi, N., Zeroual, A., & Auhmani, K. (2015). Gene-expression-based cancer classification through feature selection with KNN and SVM classifiers. 2015 Intelligent Systems and Computer Vision, ISCV 2015. doi:10.1109/ISACV.2015.7106168
[5] Cho, S.-B., & Won, H.-H. (2003). Machine learning in DNA microarray analysis for cancer classification. Proceedings of the First Asia-Pacific Bioinformatics Conference on Bioinformatics 2003-Volume 19, (May 2014), 189–198.
[6] Devika, R., Avilala, S. V., & Subramaniyaswamy, V. (2019). Comparative study of classifier for chronic kidney disease prediction using naive bayes, KNN and random forest. Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019, (Iccmc), 679–684. doi:10.1109/ICCMC.2019.8819654
[7] Ghosh, P., & Hasan, Z. (2020). A Comparative Study of Machine Learning Approaches on Dataset to A Comparative Study of Machine Learning Approaches on Dataset to Predicting Cancer Outcome, (January).
[8] Günaydin, Ö., Günay, M., & Şengel, Ö. (2019). Comparison of lung cancer detection algorithms. 2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT 2019. doi:10.1109/EBBT.2019.8741826
[9] Gupta, M., Sharma, S. K., Saxena, R., & Arora, S. (2021). Analysis of machine learning algorithms in brain tumour prediction. Journal of Physics: Conference Series, 2070(1), 012090. doi:10.1088/1742-6596/2070/1/012090
[10] Havaei, M., Jodoin, P. M., & Larochelle, H. (2014). Efficient interactive brain tumor segmentation as within-brain kNN classification. Proceedings - International Conference on Pattern Recognition, 556–561. doi:10.1109/ICPR.2014.106
[11] Ibrahim, I., & Abdulazeez, A. (2021). The Role of Machine Learning Algorithms for Diagnosing Diseases. Journal of Applied Science and Technology Trends, 2(01), 10–19. doi:10.38094/jastt20179
[12] Kalaiyarasi, M., Dhanasekar, R., Sakthiya Ram, S., & Vaishnavi, P. (2020). Classification of Benign or Malignant Tumor Using Machine Learning. IOP Conference Series: Materials Science and Engineering, 995(1). doi:10.1088/1757-899X/995/1/012028
[13] Kumar, A., Sushil, R., & Tiwari, A. K. (2019). Machine Learning Based Approaches for Cancer Prediction: A Survey. SSRN Electronic Journal. doi:10.2139/ssrn.3350294
[14] Laghmati, S., Tmiri, A., & Cherradi, B. (2019). Machine learning based system for prediction of breast cancer severity. Proceedings - 2019 International Conference on Wireless Networks and Mobile Communications, WINCOM 2019, 1–5. doi:10.1109/WINCOM47513.2019.8942575
[15] Li, R., & Sun, Y. (2010). Diagnosis of breast tumor using SVM-KNN classifier. Proceedings - 2010 2nd WRI Global Congress on Intelligent Systems, GCIS 2010, 3, 95–97. doi:10.1109/GCIS.2010.278
[16] Neagu, D. C., Guo, G., Trundle, P. R., & Cronin, M. T. D. (2007). A comparative study of machine learning algorithms applied to predictive toxicology data mining. ATLA Alternatives to Laboratory Animals, 35(1), 25–32. doi:10.1177/026119290703500119
[17] Osareh, A., & Shadgar, B. (2010). Machine learning techniques to diagnose breast cancer. 2010 5th International Symposium on Health Informatics and Bioinformatics, HIBIT 2010, 114–120. doi:10.1109/HIBIT.2010.5478895
[18] Sachin, P., & Kumar, R. Y. (2017). Detection and Classification of Blood Cancer from Microscopic Cell Images Using SVM KNN and NN Classifier. International Journal of Advance Research, 3(6), 315–324. Retrieved from www.ijariit.com
[19] Saturi, R., Scholar, R., Sai Phani, K. V, & Chand, P. P. P. (2021). A FRAME WORK TO DETECT BREAST CANCER USING KNN and SVM. European Journal of Molecular & Clinical Medicine, 08(03), 2021.
[20] Sharma, K., Kaur, A., & Gujral, S. (2014). Paper ID 1411003B , 15–20.
[21] Tahmooresi, M., Afshar, A., Bashari Rad, B., Nowshath, K. B., & Bamiah, M. A. (2018). Early detection of breast cancer using machine learning techniques. Journal of Telecommunication, Electronic and Computer Engineering, 10(3–2), 21–27.
[22] Wasule, V., & Sonar, P. (2017). Classification of brain MRI using SVM and KNN classifier. Proceedings of 2017 3rd IEEE International Conference on Sensing, Signal Processing and Security, ICSSS 2017, 218–223. doi:10.1109/SSPS.2017.8071594



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)