Table of Contents

**Chapter 11**

Introduction1

Scope1

Aim2

Objective2

**Chapter 23**

Literature Review3

**Chapter 311**

Experimental Setup11

Methodology11

Problem Understating11

Data11

Exploratory Data Analysis13

Models or Algorithm14

Data Visualization 14

**Chapter 4………………………………………………………………………………………………………………………………………….16**

Technology……………………………………………………………………………………………………………………………………….16

Deployment Phase……………………………………………………………………………………………………………………………16

**Chapter 5………………………………………………………………………………………………………………………………………….20**

Results………………………………………………………………………………………………………………………………………………20

Conclusion………………………………………………………………………………………………………………………………………..21

**References22**

**Plagiarism report………………………………………………………………………………………………………………………………24**

**CHAPTER-1**

**1.1 INTRODUCTION:**

A lot has been said in history about how precision medicine and, more precisely, how genetic testing goes to disrupt the way diseases like cancer are treated. But often only partially it happens because of the large and tedious amount of manual work still required. The real workflow goes as once sequenced; a cancer tumor can have thousands of genetic mutations [5]. But the real challenge is to bifurcate the mutations that contribute to tumor growth (drivers) from the neutral mutations (passengers). In the present time, this interpretation of genetic mutations is being done manually which is often a time-consuming task where a clinical pathologist/molecular pathologist manually reviews and classifies every single chromosomal mutation-supported evidence from text-based clinical literature [6].

DATA-🡪 Memorial Sloan Kettering Cancer Centre (MSKCC)

**PROBLEM STATEMENT:**

Classify the given genetic variations/mutations based on proof from text-based clinical writings.

**1.2 SCOPE:**

What happens in the real world right now?

* The workflow is as follows

1.A molecular pathologist chooses a list of genetic variations of interest that he/she needs to break down.

2.The molecular pathologist looks for proof in the clinical writing that in some way or another is pertinent to the genetic variations of interest

3.Lastly, this molecular pathologist invests a lot of time analyzing the proves related to each of the variations to characterize them

* **The purpose of ML in this context is to speed up the analysis that research that has been collected to determine which of the classes it’s belongs**

**1.3 AIM**

* Our objective here is to replace stage 3 with a machine learning model. The atomic pathologist will in any case need to choose which variations are of interest, and furthermore gather the important proof for them. However, the last stage, which is also the most time-consuming and tedious, would be fully automated.

**1.4 OBJECTIVE:**

* The main objective is to assess the correctness in classifying data with respect to the efficiency and effectiveness of each algorithm in terms of features selection, data cleaning, confusion matrix, accuracy, and precision.

**CHAPTER-2**

**2.1 LITERATURE REVIEW:**

The following are the reviews made by each of our team members from the research papers they read.

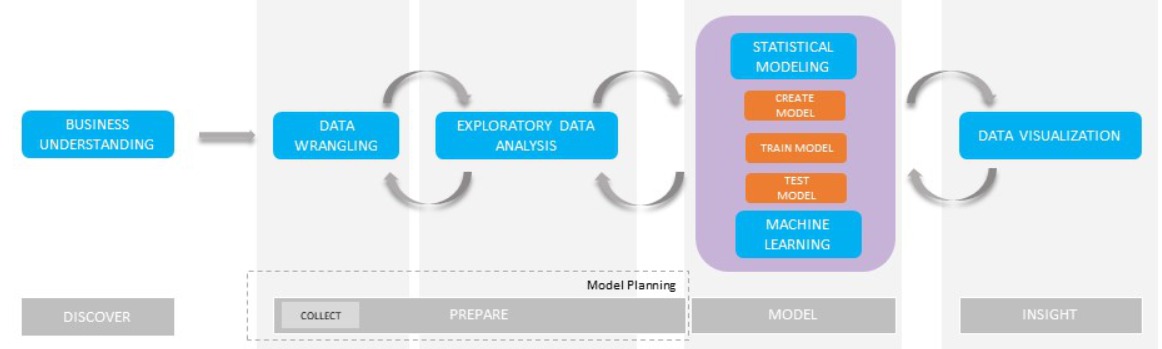
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SR.NO | PAPER TITLE | AUTHOR’S NAME | YEAR OF PUBLICATION | METHODOLOGY | ADVANTAGES | LIMITATIONS |
| 1. | The Future of Health care: Machine Learning | M.A.Jabbar , Shirina Samreen , Rajanikant  Aluvalu | 2018 | Factual strategies have been utilized to separate certain information, however measurable examination requires numerical foundation. Factual investigation is tedious as the analyzer needs to detail and test every speculation, though ML robotizes the age and testing of a theory. | From this paper, we found out about the general significance of AI in the medical care and furthermore how enormous tech goliaths like Enlitic, MedAware, and Google, have dispatched gigantic undertakings zeroed in on improving ML and computerized reasoning frameworks for the medical services framework | With this blast in advancement, it's significant for medical services experts and different partners to comprehend the guidelines set up for the viable turn of events and organization of these innovations.  Lastly, I can state that,"We can't simply put out these utilizations of man-made brainpower without getting endorsement from administrative specialists," |
| 2. | Statistical analysis of cancer data | Dr.M.Marimuthu, V. Keerthika , Assistant Professors  C.P. Sri Chidambaram S. Sreenath | 2015 | Commonness addresses new and prior cases alive on a specific date, as opposed to frequency, which reflects new instances of a condition analyzed during a given timeframe. | In this paper, we found out about the diverse measurable ways to deal with manage various sorts of highlights. In this, the strategies utilized were speculation trying like Examination of Fluctuation, relapse investigation | factual programming that figures total commonness dependent on restricted span predominance measurements. Complete commonness addresses the extent of individuals alive on a specific day who recently had a conclusion of an illness, paying little mind to how some time in the past the sickness (i.e., malignancy) was analyzed. |
| 3. | Preprocessing Breast Cancer Data to Improve the Data Quality and Diagnosiss procedure | Zeinab Sajjadnia ,  Raof Khayami  and Mohammad Reza Moosavi | May 27, 2020 | To distinguish the potential mistakes in each field, first, the information type, the area, the adequate qualities, the worth reach, and the worth configuration of each property were recognized independently | In this paper, we had the chance to get familiar with the strategies utilized for information preprocessing and Numerical factual plan which helps in the process like exactness, review, affectability, G-Mean, and F-measure in incredible detail. The dataset utilized was the RROC bosom disease dataset | In contemplating malignancy, there is a wide assortment of highlights, some of which may exist between various datasets. For instance, among the datasets of the Reza Radiation Oncology Center (RROC), there are highlights that have not been considered in any of the benchmark dataset |
| 4. | Random forest classifier combined with feature selection for breast cancer diagnosis and Prognostic | Cuong Nguyen , Yong Wang, Ha Nam Nguyen | May 2013 | This sort of survey is helpful in assembling a volume of writing in a particular branch of knowledge and summing up and incorporating it. Story Surveys are pointed toward distinguishing and summing up what have been recently distributed, keeping away from duplications, and looking for new investigation zones not yet tended to | In this paper we had the opportunity to find out around one of the troupe models know as a choice tree and how it very well may be utilized in bosom malignancy analysis. The arbitrary woods (RF)  Calculations structure a group of arrangement strategy that depends on the mix of a few choice trees | This examination has a few limits. This investigation has a set number of articles inspected because of few watchwords and data sets utilized. Likewise, these examinations just talk about the maternal and neonatal reference framework from patient and medical care framework points of view. |
| 5. | Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis | Hiba Asria ,Hajar Mousannifb ,Hassan Al Moatassime c ,Thomas Noeld | 3 October 2018 | The destructive cells are named Kind (B) or Dangerous (M). There are numerous calculations for the grouping and forecast of bosom malignancy: Backing Vector Machine (SVM), Choice Tree (Truck), Gullible Bayes (NB), and k Closest Neighbors (KNN). In this undertaking, the Help Vector Machine (SVM) on the Wisconsin Bosom Disease dataset is utilized. | This paper is about bosom malignant growth for ladies and predicts the danger of getting Bosom disease to them. They work on the Wisconsin Bosom malignant growth dataset the primary Focal point of the creator was which calculation abuses better viability, Which calculation more proficient? Which calculation gives higher exactness? They assessed the viability dependent on Kappa Measurement, Mean Absolute Blunder, Root Mean Square. They utilize a 10-crease Cross-Validation Test | It doesn't give the legitimate thought of Which calculation abuses better adequacy? Which calculation is more proficient? Which calculation gives a higher precision? |
| 6. | Personalized cancer therapy | The Sheikh Khalifa Bin Zayed Al Nahyan 1Department of Investigational Cancer Therapeutics, The University of Texas, MD Anderson Cancer Center, Houston, Texas 77030, USA;2The Sheikh Khalifa Bin Zayed Al Nahyan Institute for Personalized Cancer Therapy, The University of Texas, MD Anderson Cancer Center, Houston, Texas 77030, USA;3Department of Breast Surgical Oncology, The University of Texas, MD Anderson Cancer Center, Houston, Texas 77030, USA1Department of Investigational Cancer Therapeutics, The University of Texas, MD Anderson Cancer Center, Houston, Texas 77030, USA;2The Sheikh Khalifa Bin Zayed Al Nahyan Institute for Personalized Cancer Therapy, The University of Texas, MD Anderson Cancer Center, Houston, Texas 77030, USA;3Department of Breast Surgical Oncology, The University of Texas, MD Anderson Cancer Center, Houston, Texas 77030, USA | 2018 April | Choosing the correct therapy for patients with malignancy is an unpredictable choice dependent on constantly advancing sub-atomic diagnostics and quickly arising biomedical writing. Following relationship between significant genomic adjustments and focused on treatments in clinical preliminaries | From this paper, I finish up and got information base which is at present dependent on DNA sequencing, joining of RNA and protein examination may give extra data on tumor science and further treatment choices later on | The data introduced on this site isn't expected to give direct therapy suggestions but instead to integrate what is thought about malignant growth related qualities and their suggestions for disease treatment |
| 7. | Cancer classification and prediction using logistic regression | Xiabozhou, Kuang Yuo-lie, Stephan T.C.Wong | 11 september,  2004 | These record profiling methods have been applied to contemplate the examples of quality articulation across numerous tests that overview a wide assortment of cell reactions, aggregates, conditions, and regularly through perceptions at different time | From this paper, I infer that sensible relapse will be utilized to distinguish significant qualities, a similar calculated relapse model is then utilized for disease characterization and expectation | It is seen quality 10 and quality 336remain vital for various commotion levels. The outcomes demonstrate that the proposed techniques are not delicate to the diverse commotion levels. |
| 8. | Support vector machine for cancer diagnosisis | Nasser H.Sweilam , A.A.Tharwat , M.K.Abdul Monium | DECEMBER 2010 | Approach for Malignant growth Analysis Model by pre-handling the information utilizing scaling (we scale straightly each quality to the scope of [0, 1]), pre-prepared information are part into preparing and testing (free) datasets. | From this paper, I reason that malignancy can be harmful and non-dangerous. Where destructive tissues infiltrate and annihilate the sound body tissues and along these lines to order this disease we can build a nonlinear ideal classifier utilizing a help vector machine | When utilizing the SVM, three impediments are gone up against: how to pick the part work and ideal information include subset for SVM, and how to set the best portion boundaries. These hindrances are significant on the grounds that the element subset decision impacts the proper piece boundaries and the other way around |
| 9. | CANCER DIAGNOSIS USINGNAIVE BAYESALGORITHM | Rashmi M, Usha K Patil | May 2016 | Quality information from the malignant growth patients will be put away in the capacity worker and for the new tolerant; we do the important tests and will get the qualities articulation esteems, in light of these qualities’ framework will order the kind of the disease. Information mining innovation helps in characterizing disease patients and this strategy assists with distinguishing potential malignant growth patients by just examining the information | From this paper, I presume that Quality information from the disease patients will be put away in the capacity worker and for the new persistent we do the essential tests and will get the qualities articulation esteems, in view of these qualities’ framework will order the sort of the malignancy | Manual interaction is delayed as after the radiologist's audit the specialist needs to survey likewise and give his/her comments lastly tell if the disease is available or not. There is a need to computerize this cycle to make the malignant growth finding productive and quick with the utilization of best-in-class innovation. |
| 10. | Cancer Detection Using K-Nearest Neighbour Algorithm | Shagun Chawla, Rajat Kumar, Ekaansh Agrawal, Sarthak Swain | 2018 | We have executed K-Closest Neighbors Calculation utilizing different standardization procedures and distance capacities at various estimations of K. A similar report utilizing different standardization procedures, i.e., Min-Max standardization, Z-Score standardization and Decimal Scaling standardization, and diverse distance measurements | From this paper, I came to realize that Even with the advancement of new innovations in the field of medication and examination, the exact conclusion of this lethal illness result is quite possibly the main undertakings should have been done to date. Our goal is to build up a refined and computerized analytic framework that yields precise and reproducible outcomes for foreseeing whether a bosom disease tumor is kind (non-harmful) or dangerous (carcinogenic) | Bosom malignancy is extremely well known among females everywhere on the world. Be that as it may, identifying this malignancy in its first stages helps in quite a while. Radiologists can anticipate if the mammography pictures have malignancy or not, however they may miss about 15% of them. |
| 11. | Is Combining Classifiers with Stacking Better than Selecting the Best One | Bernard Zenko , Saso Dzeroski | 2004 | We propose two augmentations of this technique, one utilizing an all-inclusive arrangement of meta-level highlights and the other utilizing multi-reaction model trees to learn at the meta-level. We show that the last expansion performs better compared to existing stacking approaches and better than choosing the best classifier by cross-validation | We have observationally assessed a few cutting-edge strategies for developing outfits of heterogeneous classiﬁers with stacking and shown that they perform, (best case scenario, similarly to Choosing the best classiﬁer from the group by cross-validation. | The main issues in stacking are presumably the decision of the highlights and the calculation for learning at the meta-level. Stacking with PDs and MLRs tends to both |
| 12. | Machine Learning Classification Techniques for Breast Cancer Diagnosis | David A. Omondiagbe , Shanmugam Veeramani , Amandeep S. Sidhu | 2018 | This paper targets exploring Backing Vector Machine (utilizing spiral premise portion), Fake Neural Organizations, and Credulous Bayes utilizing the Wisconsin Demonstrative Bosom Disease (WDBC) Dataset. | This paper dissected the WDBC dataset utilizing dimensionality decrease strategies and three famous MLalgorithms to characterize harmful and kindhearted tumors. The exploratory work demonstrates that characterization execution is subject to the ML arrangement strategy picked | Highlight measurements can be decreased utilizing the suitable component choice or highlight extraction technique. There are a few strategies used to decrease the components of highlights in a dataset |
| 13. | Applications of Machine Learning in Cancer Prediction and Prognosis | Joseph A. Cruz, David S. Wishart | 2014 | The principal objective of ML methods is to create a model which can be utilized to perform characterization, expectation, assessment, or some other comparative assignment. The most well-known errand in the learning interaction is the grouping | In this audit, we have endeavored to clarify, look at and evaluate the presentation of various ML that are being applied to disease expectation and guess. | Notwithstanding, the utilization of highlight determination procedures may bring about explicit vacillations concerning the formation of prescient component records. A few investigations in the writing talk about the marvel of absence of arrangement between the prescient quality records found by various gatherings |
| 14. | ML | Cancer cell classification using Scikit-learn | Rahul Roy | 2018 | characterize disease cells dependent on their highlights, and distinguishing on the off chance that they are 'threatening' or 'kind'. We will utilize Scikit-learn for an AI issue. | This is the nearest approach taking all things together the exploration papers managing grouping errands. In the paper, we are ordering two kinds of cell and in our work, we are grouping 7 sorts of variety in qualities | In. It has a monstrous issue  With programmed AI and profound learning pipelines. |
| 15. | Clinical Data Extraction and Normalization of Cyrillic Electronic Health Records Via Deep-Learning Natural Language Processing | Boyang Zhao, PhD | 2019 | In this paper they have utilized double embeddings for English and Bulgarian dialects, encoding both syntactic and extremity data for the words. The embeddings were consequently adjusted all together that they were inside a similar vector space. The embeddings were utilized as a contribution to convolutional or intermittent neural organizations to infer the biomarker status of estrogen receptor, progesterone receptor, and human epidermal development factor receptor 2. | They can resolve vagueness is exceptionally factor clinical content containing both Latin and Cyrillic content. Last models joining both English and Bulgarian punctuation and extremity embeddings accomplished F1 scores of 0.90 or higher for all estrogen receptor, progesterone receptor, and human epidermal development factor receptor 2 biomarkers. The models were hearty against human mistakes initially found in the preparation set. Additionally, such models can be stretched out for dissecting text containing words not seen during preparing. | They can resolve equivocalness is profoundly This paper manage the content information Identified with bosom malignant growth and we are motivated by the methodology of utilizing text information utilizing conv1d yet this paper, it just arrangements with the information of bosom disease and we are taking content information of malignancy transformation and attempting distinctive design to improve precision |

**CHAPTER-3**

**3.1 EXPERIMENTAL SETUP:**

We are going to train our model on Intel(R) Core (TM) i7-8550U CPU @ 1.80GHz, 8Gb RAM,64 -bit OS, Windows 10 by Google Colab, and Jupyter Notebook(Anaconda 3) (For the offline purpose).

**3.1.1 METHEDOLOGY (LIFE CYCLE OF OUR PROJECT):**



1) Problem Understating:

In this step, we are analyzing the problem and coming with a solution that can help the user so we come up with these constraints

* No low-latency requirement.
* Interpretability is important.
* Errors can be very costly.
* The probability of a data-point belonging to each class is needed.

2) Data

2.1) Data Overview:

* Source: <https://www.kaggle.com/c/msk-redefining-cancer-treatment/data>
* We have two data files: one contains the information about the genetic mutations and the other contains the clinical proofs (text) that doctors/pathologists use to classify the genetic mutations.
* Both these data files are having a common column called ID
* Data file's information:
  + 1. training\_variants (ID, Gene, Variations, Class)
    2. training\_text (ID, Text)

2.2) Example of Data point



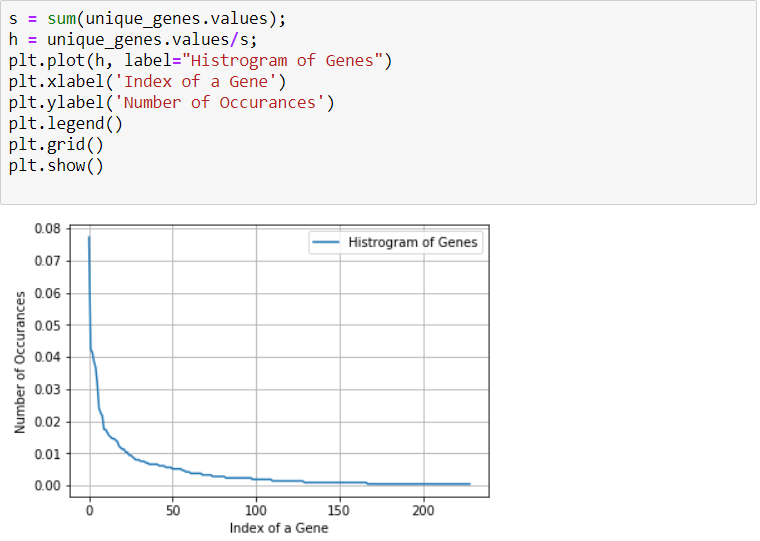
3) Exploratory Data Analysis

In this step, we will understand data and find some meaningful insight

It requires:

1. Variable Identification
2. Univariate Analysis
3. Bi-variate Analysis
4. Missing values treatment
5. Outlier treatment
6. Variable transformation
7. Variable creation

Univariate Analysis GENE data



4) Models or Algorithm

In this part, we will perform different types of ML algorithm to train our model for solving our Problem Statement based on the performance metric and the data we will check the accuracy of the model.

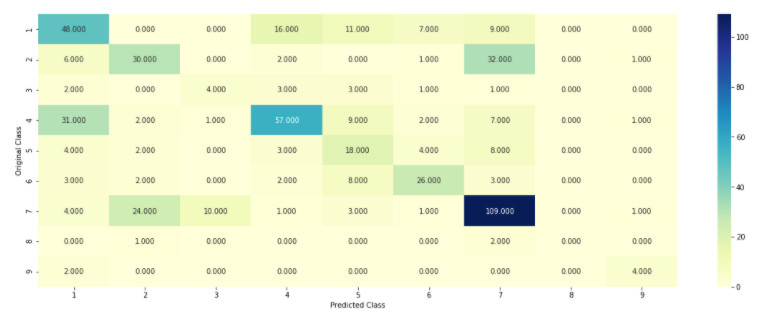
It consists of the following things:

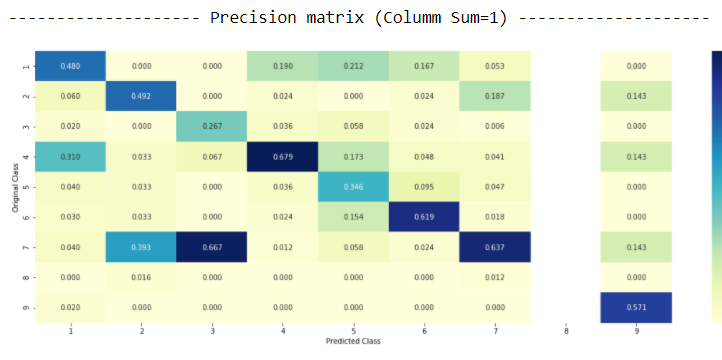
1. Naive Bayes
2. KNN
3. Logistic Regression
4. SVM
5. Random Forest
6. Stalking Model (Maximum Voting Classifier)

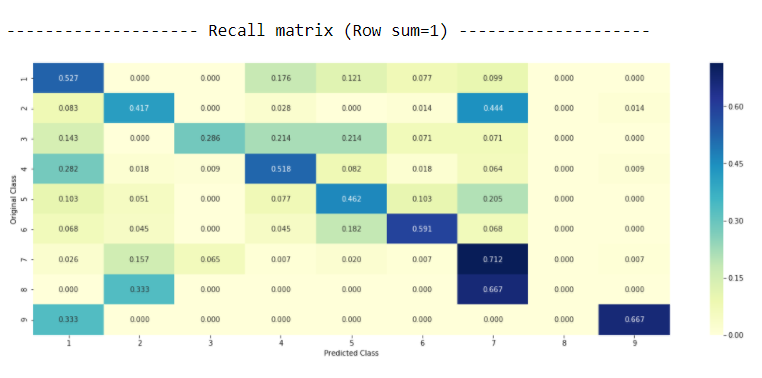
5) Data Visualization

In this step, we visualize the predictions of models and results of our performance metrics [7] so it will be easy to see the behaviour of the model so if there is any error, we can repeat the above step for correction Visualization to check by Performance Metrics

Confusion Matrix







**CHAPTER-4**

**4.1 TECHNOLOGY:**

To make a User Interface we have used Tkinter.

Based on extensive research and discussions with our guide we concluded that for training and testing we use various models like SVM, Naive Bayes, KNN, Logistic Regression, Random Forest, and Stalking Model (Maximum Voting Classifier) and amongst all which was the most optimum was used.

**4.2 DEPLOYMENT PHASE:**

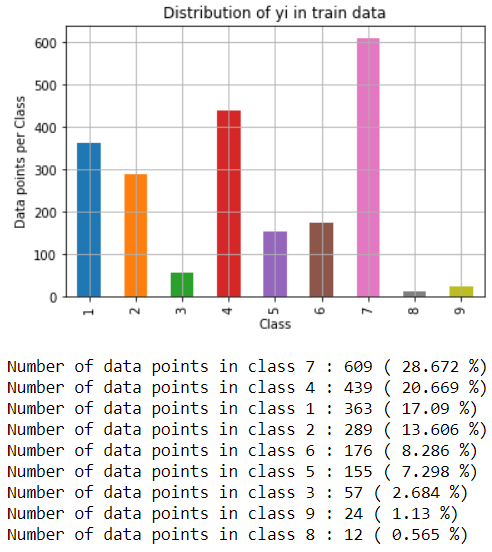
Mapping the real-world problem into an ML problem, the first thing was to understand the pattern and the data labels and then majorly the big task was to find which type of machine learning would work best at the initial stage and for our problem, There are nine different classes genetic mutation can be classified into => Multi-class classification problem

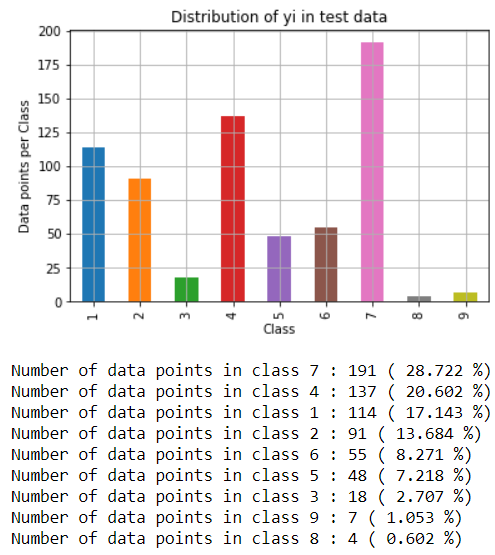
Then the backbone for the performance that is performance matrix was taken into account so that based on that we would find the suitable model also could make changes accordingly, so according to our criteria that is the probability of data points should be known and also the interpretability to the doctor who is seeing the result we concluded that Multi-class log-loss, Confusion matrix

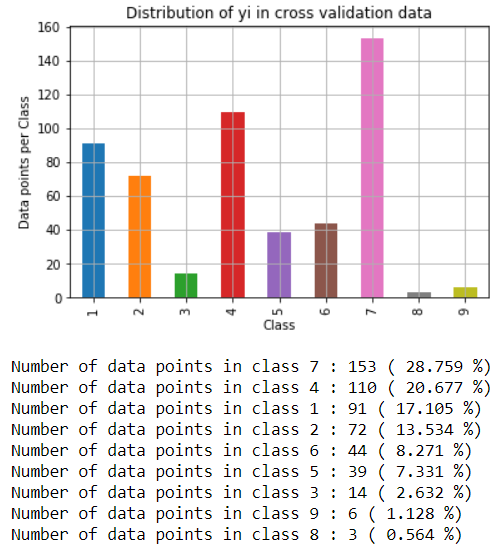
Now, comes the main data phase in which we have to train, cross-validate, and test the data set. So based on the research papers and also with the help of our guide we have split the dataset randomly into three parts train, cross-validation, and test with 64%,16%, 20% of data respectively

After splitting the data, our first step was to check the labeled distribution in the train, cv, and test data, so that this helps the model to avoid overfitting and underfitting. After that by the help of python libraries like pandas, Matplotlib, NumPy, Sklearn, SciPy and many other which were suitable to explore the data in depth has been used to make any conclusions for the further steps. The exploratory data analysis, preprocessing textual data and other related things are explained in the previous chapter.

After seeing the distribution of data, it was clear that training data, test data, and cross-validation data have the same distribution of labeled data which helped us to avoid a scenario like overfitting and underfitting and also by this we were sure that we do not have to apply techniques such as data oversampling and data oversampling and other related techniques to equalize the label distribution.



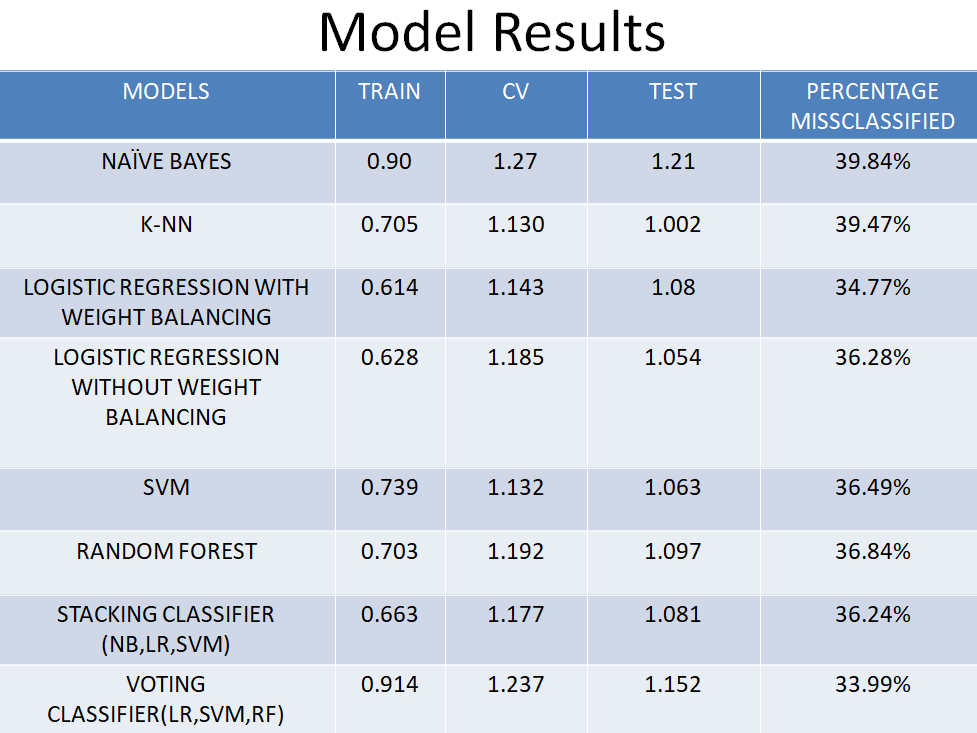




We have chosen the appropriate featurization based on the ML model we use. For this problem of multi-class classification with categorical features, one-hot encoding is better for Logistic regression while response coding is better for Random Forests.

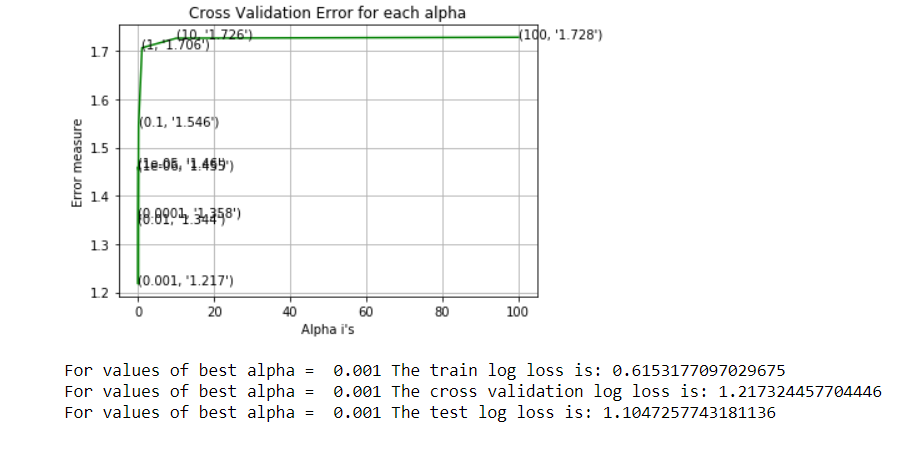
Then we created a Random model in which we read the data and just gave the prediction and a log-loss value based on guessing, that we had to write that value as a boundary case and we had to find the model which has the log-loss value less than that threshold value and the model which has the lowest value in comparison to the random model we would choose that as the final model. By this approach, we started our algorithm building journey

After applying the following models by the above approach based on past research papers in the medical domain and all other sources like blogs and discussion portals online, we concluded that the following models would suit best, and also for a better understanding, we have depicted the result of each model in a tabular format.



By this it stages it was clear to use Logistic Regression with weight balance was the best amongst all because, as it has very less misclassification percentage, it is interpretable, also minor class gets good preferences.

The following is the hyper parameter turning

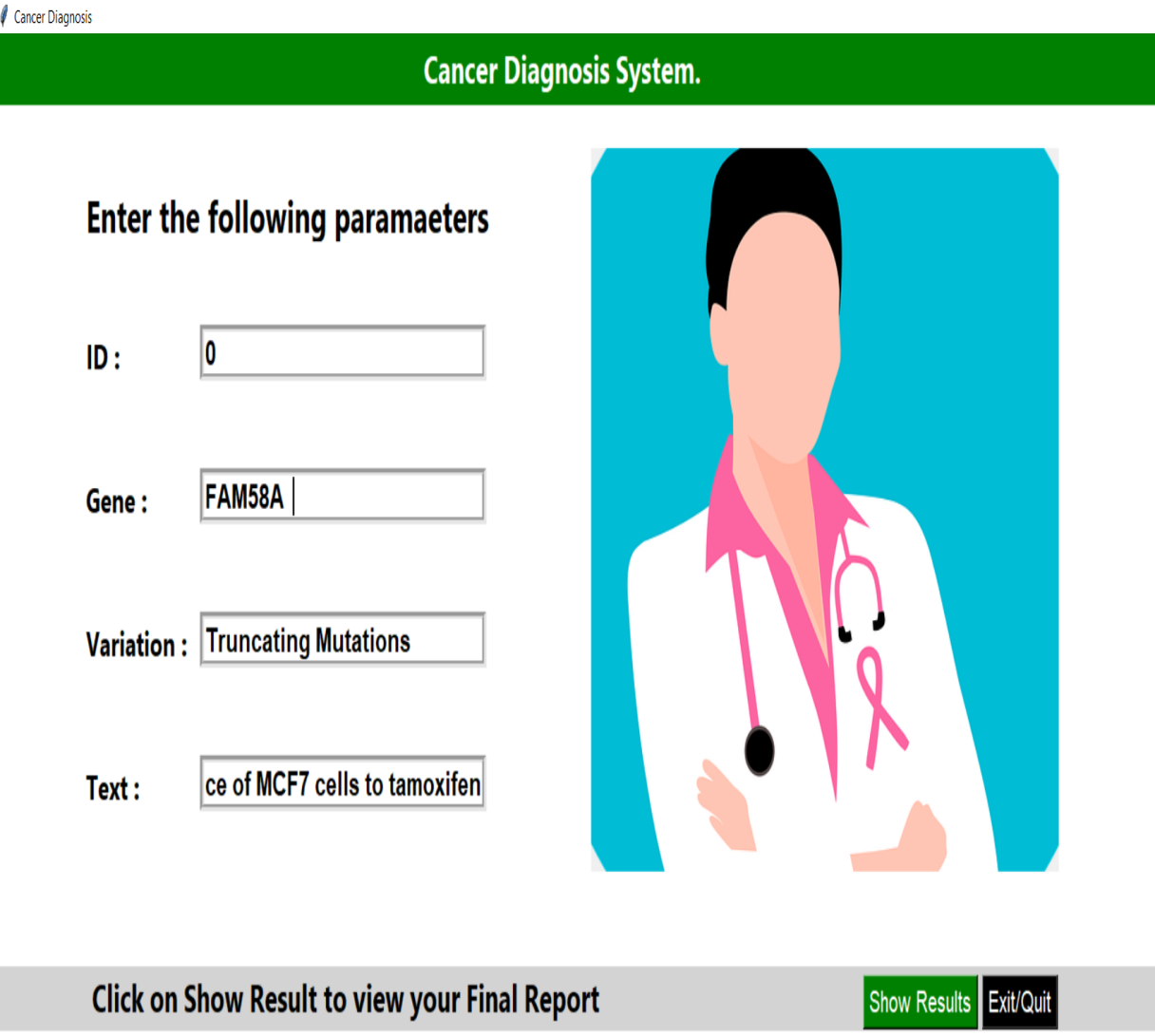


**CHAPTER-5**

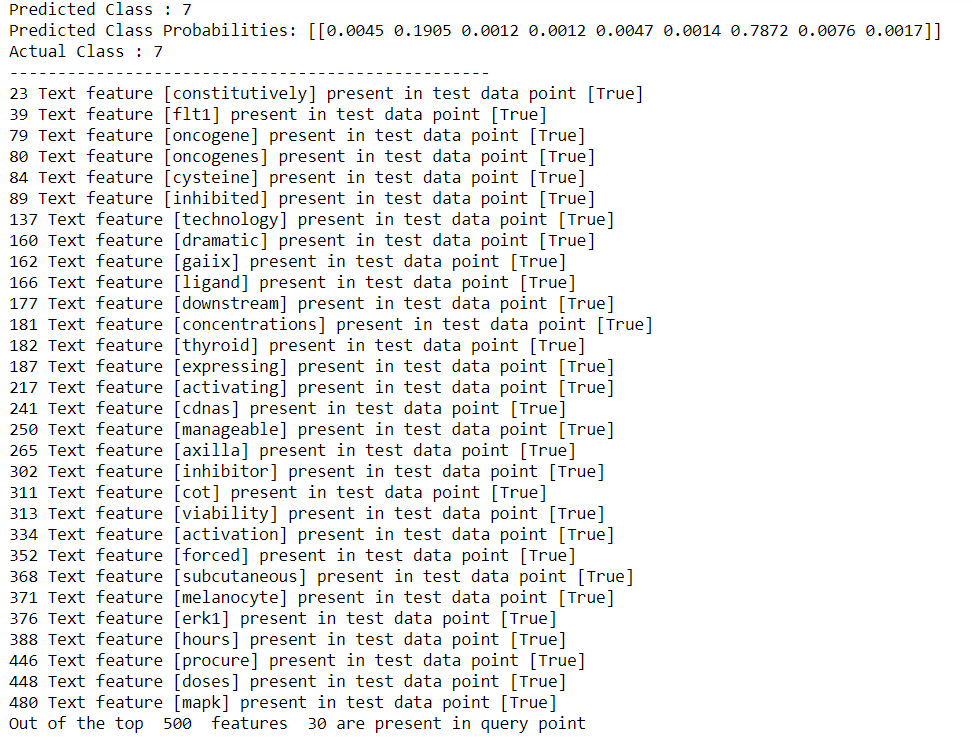
**5.1 RESULTS:**

The user- interface was made using tkinter as it was compatible and very smooth with our models, here are the input and output of our final model

**INPUT:**

****

**OUTPUT:**

****

**CONCLUSION:**

In today’s scenario the workflow goes like this, a molecular pathologist selects a list of genetic variations of interest that he/she wants to analyze. The molecular pathologist searches for proofs in the medical writings that somehow are relevant to the genetic variations of interest. Finally, this molecular pathologist invests a lot of time analyzing the evidence related to each of the variations to classify them. So, by applying our ML project in this context, would speed up the analysis that research has been collected to determine which of the classes belongs to the doctors and would save a lot more lives as compared to recent times.

**REFRENCES:**

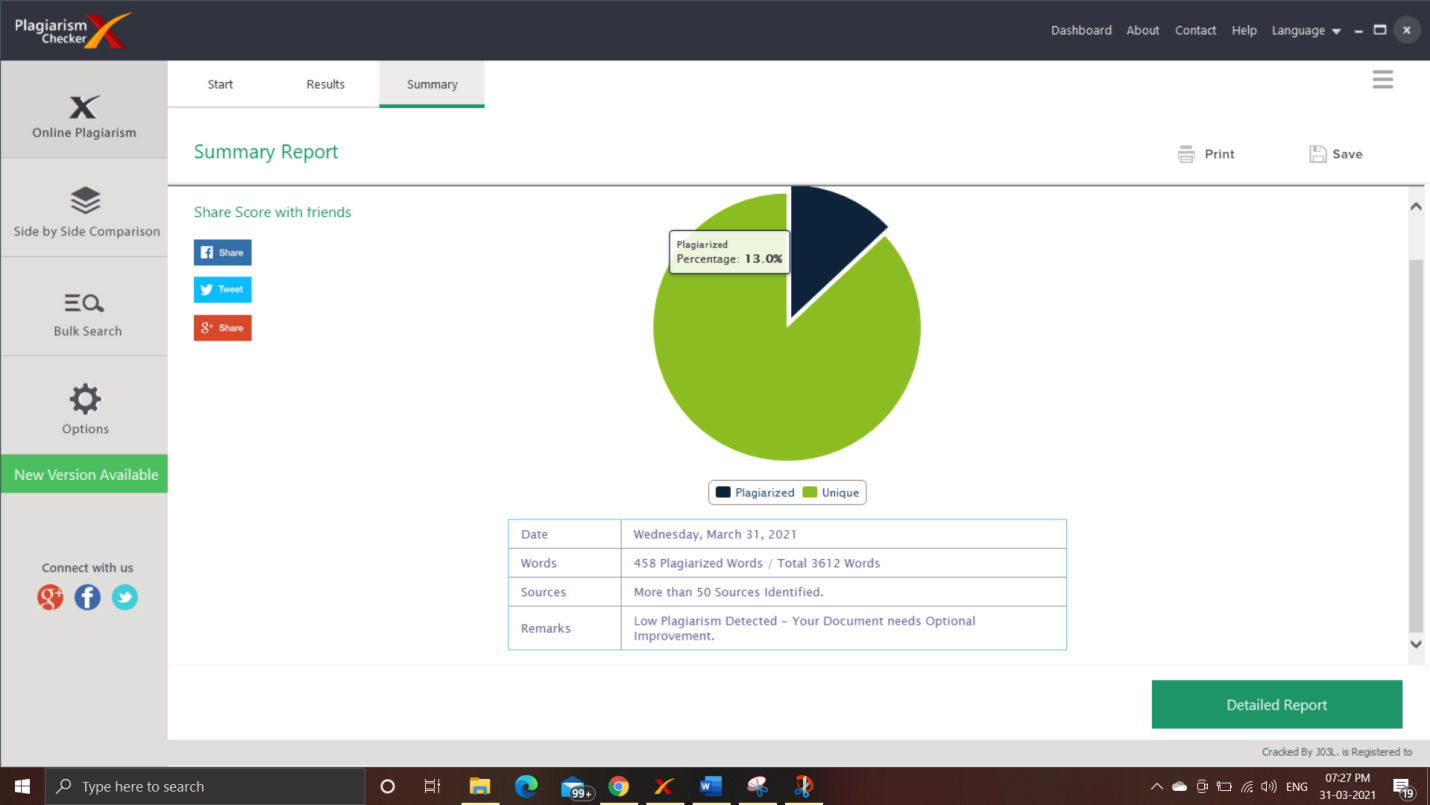
1. Predicting breast cancer recurrence using data mining techniques., IEEE,Fan Q , Zhu CJ, Yin L. ,2010 International Conference on Bioinformatics and Biomedical Technology. New York 2010:310-311.
2. “UCI Machine Learning Repository: Breast Cancer Wisconsin (Original) Data Set.” [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29. [Accessed: 29-Dec-2015].
3. Kallenberg, M.; Petersen, K.; Nielsen, M.; Ng, A.Y.; Diao, P.; Igel, C.; Vachon, C.M.; Holland, K.; Winkel, R.R.; Karssemeijer, N.; et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Trans. Med. Imaging 2016, 35, 1322–1331
4. Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis, [Computational and Structural Biotechnology Journal](https://www.sciencedirect.com/science/journal/20010370),, Hiba Asria ,Hajar Mousannifb ,Hassan Al Moatassime ,Thomas Noeld, The 6th International Symposium on Frontiers in Ambient and Mobile Systems (FAMS 2016),2016, Morocco
5. Statistical Analysis of Cancer Data, International Conference on Mathematical Modelling and Scientific Computing( ICMMSC - 2018) @ IIT INDORE, Dr. M. Marimuthu , V. Keerthika , Assistant Professors C.P. Sri Chidambaram S. Sreenath , R. Mohan Kumar , MSC Data Science Department of computing Coimbatore Institute of Technology , 2018, INDIAN INSTITUTE OF TECHNOLOGY, INDORE
6. Multivariate Analysis of Breast Cancer Prediction Parameters  Proceedings of the 1st Annual World Conference of the Society for Industrial and Systems Engineering, Benjamin Schleich, Hema Sudarsanam, and Chanchal Saha , Department of Systems Science and Industrial EngineeringState University of New York at Binghamton, 2012, Washington, D.C.
7. A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications,Healthcare(basel),[Muhammet Fatih Ak](https://www.ncbi.nlm.nih.gov/pubmed/?term=Ak%20MF%5BAuthor%5D&cauthor=true&cauthor_uid=32357391), PMC,2020
8. Comparison of Data Mining Algorithms for Predicting the Cancer Disease Using Python, IEEE , [Mehtab Mehdi](https://ieeexplore.ieee.org/author/37088422342),[Kanika Pahwa](https://ieeexplore.ieee.org/author/37086301543),[Bharti Sharma](https://ieeexplore.ieee.org/author/37088423028),2020,Moradabad, India,
9. "Supervised machine learning: A review of classification techniques", IOS Press, SB Kotsiantis and I Zaharakis, in Emerging Artificial Intelligence Applications in Computer Engineering, 2007.
10. Microarray-based cancer prediction using single genes, Wang, X., Simon, R. *BMC Bioinformatics* 12, 391 (2011)

# Breast Cancer Diagnosis using Python, EasyChair Preprint, Salma Yousaf, Shakeel Khan, June 2020, Department of Computer Science, Riphah International University Lahore, Pakistan

# Using three machine learning techniques for predicting breast cancer recurrence , Ahmad, L. G., Eshlaghy, A. T., Poorebrahimi, A., Ebrahimi, M., & Razavi, A. R. (2013).. J Health Med Inform, 4(124), 3.

# Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science, 83, 1064-1069

**Plagiarism report:**

****