

TOPIC: DIETARY PATTERN EXPLORATION IN HEALTHCARE

Student details: Sree Nandini Rasiraju (23107149) and Taranjit Kaur (22196780)

Date: 2023-08-20

1 Domain description

Healthcare includes many activities, services, and behaviors to maintain and promote health. Healthcare professionals, researchers, policymakers, and patients participate (Ha & Grizzle, 2015). Healthcare is to promote well-being, prevent and cure sickness, and assure the physical, mental, and social health of people and groups. Healthcare has several subfields and expertise. Primary care, specialty care, nursing, public health, pharmaceuticals, medical research, healthcare technology, and more are included. Vaccinations, health education, and medical treatments and procedures are included in the domain. It also manages health systems, policies, and resources. Healthcare directly affects quality of life and lifespan. It covers chronic illnesses, infectious diseases, mental health problems, maternity and child health, geriatric care, and emergency medical services. Medical science, technology, demography, illness trends, socioeconomic determinants of health, and ethics affect it.

Science, technology, and data have advanced healthcare in recent years (Rahmani et al., 2018). Digital health technology, EHRs, telemedicine, AI, and genomics have altered healthcare delivery, management, and research. The field is also emphasizing patient empowerment and individualized medicine. Healthcare has difficulties and possibilities. It addresses healthcare accessibility, cost, and equality in underprivileged groups. It also addresses growing public health issues, promotes health promotion and disease prevention, and improves healthcare delivery via multidisciplinary cooperation and evidence-based approaches. The domain promotes research and innovation to create novel treatments, therapies, and interventions that improve health outcomes and healthcare system efficiency.

2 Problem definition

Dietary patterns and their health effects is a particular problem in healthcare. Poor diets cause chronic illnesses including cardiovascular disease, diabetes, and obesity. Due to individual habits, cultural norms, socioeconomic circumstances, and genetic predispositions, evaluating food patterns and health consequences is difficult. Thus, extensive and systematic dietary pattern analysis is needed to identify critical elements that influence healthy or unhealthy eating habits and generate tailored treatments and recommendations (Cao et al., 2018).

Healthcare prediction and categorization is another problem. Machine learning and data mining can examine enormous datasets to find patterns that may predict illness incidence, severity, and therapy (Berry & Linoff, 2011). Selecting the best algorithms, feature selection, and managing missing or noisy data to make accurate predictions is difficult. Healthcare professionals and patients trust models that are interpretable and explainable.

3 Literature Review

3.1 Dietary pattern analysis

Nutrition research uses dietary pattern analysis to investigate dietary patterns and health effects (Witten et al., 2016). This method addresses the synergistic effects and interactions of various dietary components rather than specific nutrients or meals. It helps researchers to determine population-specific food habits. Dietary pattern analysis often uses factor analysis, notably PCA and cluster analysis (Al-Hassan & Fawaz, 2020). PCA finds linear combinations of dietary factors that explain the most variation. It reduces data dimensionality and identifies uncorrelated dietary elements or components (Lopes & Oliveira, 2019). The food categories or nutrients that contribute most to each of these components indicate the population's primary dietary trends. Cluster analysis, on the other hand, groups people with similar food consumption habits without assuming any underlying variables (Chen & Guestrin, 2016). This method reveals population dietary groups that may represent cultural, geographical, or socio-economic disparities. Numerous researches have used dietary pattern analysis to examine food patterns and health effects (Bengio, 2019). According to studies, the Western diet is heavy in processed and fried foods, whereas the Mediterranean diet is high in fruits, vegetables, whole grains, and healthy fats (Bramer, 2016). Health consequences vary with these patterns. Western diets are linked to obesity, type 2 diabetes, cardiovascular disease, and some cancers (Alpaydin, 2020). Mediterranean diet adherents have a lower risk of chronic illnesses, better cardiovascular health, and superior health outcomes.

Dietary pattern analysis has also been used to study health issues and diet (Witten et al., 2016). Dietary habits have been linked to hypertension, metabolic syndrome, and mental illness. The Dietary Approaches to Stop Hypertension (DASH) diet, which emphasizes fruits, vegetables, whole grains, lean proteins, and low-fat dairy products, lowers blood pressure and reduces hypertension risk. A balanced diet rich in fruits, vegetables, whole grains, and omega-3 fatty acids has been linked to decreased depression risk and greater mental health (Sven, 2018). Dietary pattern analysis also helped create dietary guidelines. Researchers and policymakers may advise consumers, healthcare providers, and public health organizations by identifying good and harmful diets (Makridakis et al., 2018). These recommendations emphasize healthy diets, not only nutrients or food categories.

3.2 Machine learning and predictive models in healthcare

Healthcare research has focused on machine learning and predictive modeling because they may improve patient outcomes, illness detection and diagnosis, and healthcare delivery (Han et al., 2011). These methods extract relevant patterns and insights from massive and complicated healthcare datasets, improving forecasts and decision-making. Machine learning can anticipate and measure illness risk. Machine learning algorithms can find patterns and relationships in past patient data (Bishop, 2016). Based on demography, genetics, and lifestyle variables, these algorithms may create prediction models that estimate the chance of getting specific illnesses or ailments. Machine learning algorithms predict cardiovascular disease, diabetes, cancer, and other chronic disorders (Alpaydin, 2020). These models may help doctors identify high-risk patients and prevent illnesses.

Machine learning is used for medical image analysis and illness prediction (James, 2013). Medical imaging helps diagnose and treat illness. Medical picture interpretation is subjective and time-consuming (Goldberg, 2019). Machine learning algorithms can automate picture processing and help doctors find anomalies, cancers, and other diagnostic indicators (Rahmani et al., 2018). Convolution neural networks (CNNs) perform well in image classification, segmentation, and detection (Bramer, 2016). CNNs have been used to identify cancer, Alzheimer's, and cardiovascular problems in X-rays, CT scans, and MRIs.

In healthcare systems, machine learning algorithms optimize resource allocation, patient management, and clinical decision-making (Evans & Olson, 2014). These algorithms can examine enormous amounts of patient data, including electronic health records, test results, and vital signs, to forecast hospital readmissions, optimize bed use, and optimize drug doses (Aggarwal, 2015). Predictive models allow healthcare practitioners to intervene and create tailored treatment plans for patients, improving health outcomes and resource use (Géron, 2019). However, using machine learning in healthcare is difficult. High-quality, well-annotated

datasets are a problem. Healthcare databases are fragmented, varied, and missing values, making data extraction difficult (Berry & Linoff, 2011). Working with sensitive healthcare data requires patient privacy and data security.

3.3 Association mining and rule discovery

Association rule mining, also known as association mining, finds intriguing patterns, correlations, and linkages in massive datasets (Chiang & Wainwright, 2019). It may reveal hidden correlations and dependencies between variables, which can aid decision-making in healthcare and other fields (Aggarwal, 2015). Association mining can discover medical illnesses, symptoms, risk factors, and treatment results. Association mining algorithms may find patterns of co-occurrence or statistical relationships in patient data like electronic health records and medical claims databases. Association rules with antecedents and conclusions reflect these patterns. An association rule in healthcare may show that a genetic mutation increases the risk of a disease (Abadi et al., 2016).

Apriori algorithms are used to mine frequent itemsets and build association rules (Evans & Olson, 2014). These methods repeatedly scan the dataset to find itemsets with a minimal support threshold. The percentage of transactions that include an itemset is its support. After identifying frequent itemsets, confidence and lift may be used to build association rules. An association rule's confidence is the percentage of antecedent-containing transactions that include the consequent. High confidence values suggest significant antecedent-consequent associations (Amores et al., 2020). Lift quantifies the divergence from independence and shows how much more frequent the consequent is given the antecedent than its general occurrence in the dataset. Lift levels higher than 1 is positive.

Association mining can help healthcare applications (Han et al., 2011). For instance, it may assist healthcare practitioners establish preventative or screening programs by identifying disease risk factors. It may also help clinicians make better treatment decisions by revealing treatment-outcome relationships. Association mining can detect resource consumption, patient flow, and healthcare service utilization trends to help manage healthcare resources.

4 Data set description

Blood Pressure Abnormality - Indicates the presence of blood pressure abnormality. Possible values: 1 (Yes), 0 (No).

Level of Hemoglobin - The level of hemoglobin in the blood.

Genetic Pedigree Coefficient - The genetic pedigree coefficient of an individual.

Age - The age of the individual.

BMI - Body Mass Index (BMI) of the individual.

Sex - The gender of the individual. Possible values: 1 (Male), 0 (Female).

Pregnancy - Indicates if the individual is pregnant. Possible values: 1 (Yes), 0 (No).

Smoking - Indicates if the individual is a smoker. Possible values: 1 (Yes), 0 (No).

Physical activity - The level of physical activity of the individual.

Salt content in the diet - The salt content in the individual's diet.

Alcohol consumption per day - The amount of alcohol consumption per day.

Level of Stress - The level of stress experienced by the individual.

Chronic kidney disease - Indicates the presence of chronic kidney disease. Possible values: 1 (Yes), 0 (No).

Adrenal and thyroid disorders - Indicates the presence of adrenal and thyroid disorders. Possible values: 1 (Yes), 0 (No).

The summary gives a statistical overview of the variables. The mean prevalence of “Blood_Pressure_Abnormality” is 48.6%. “Level_of_Hemoglobin” indicates a mean of 11.62 and a moderate interquartile range. “Genetic_Pedigree_Coefficient” has a mean of 0.4929, indicating genetic lineage. “Age” has a mean age of 50.04, ranging from 25 to 75 years. “BMI” (Body Mass Index) varies from 10 to 50, averaging 30.1. 50.2% of the “Sex” variable represents gender. “Smoking” and “Physical_activity” average 51.2% and 25,211, respectively. “Salt_content_in_the_diet” averages 24,620, whereas “alcohol_consumption_per_day” averages 249.9. “Level_of_Stress” and “Chronic_kidney_disease” mean 2.018 and 49.67%, respectively. “Adrenal_and_thyroid_disorders” average 43.9%. Summary statistics reveal the dataset’s variables’ primary patterns, variability, and distributions.

5 Data set pre-processing.

Data preparation is essential to accurate and relevant data analysis. Preprocessing this dataset of health indicators is crucial to ensuring accurate and useful analysis.

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

## Loading required package: Matrix

##
## Attaching package: 'arules'

## The following object is masked from 'package:dplyr':
##
##   recode

## The following objects are masked from 'package:base':
##
##   abbreviate, write

## corrrplot 0.92 loaded

##
## Attaching package: 'xgboost'

## The following object is masked from 'package:dplyr':
##
##   slice

## randomForest 4.7-1.1

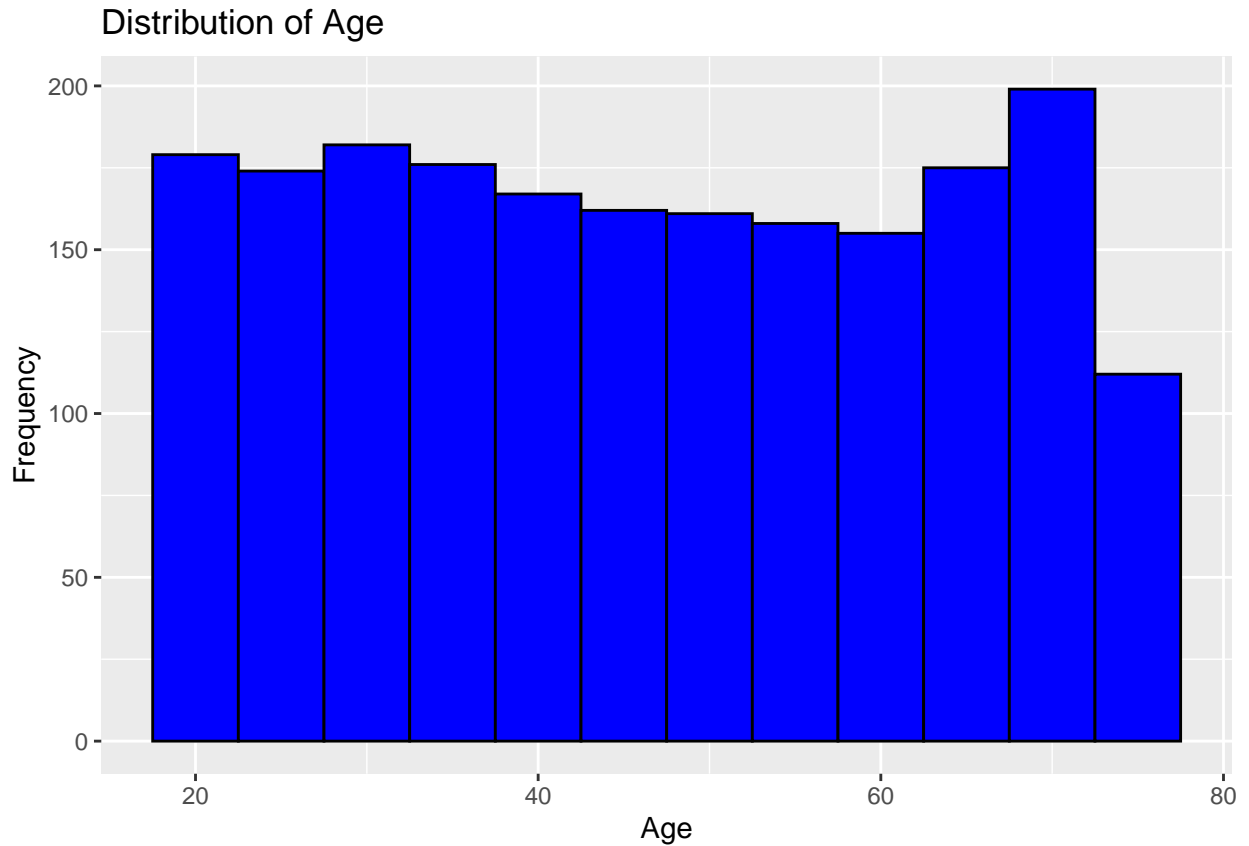
## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

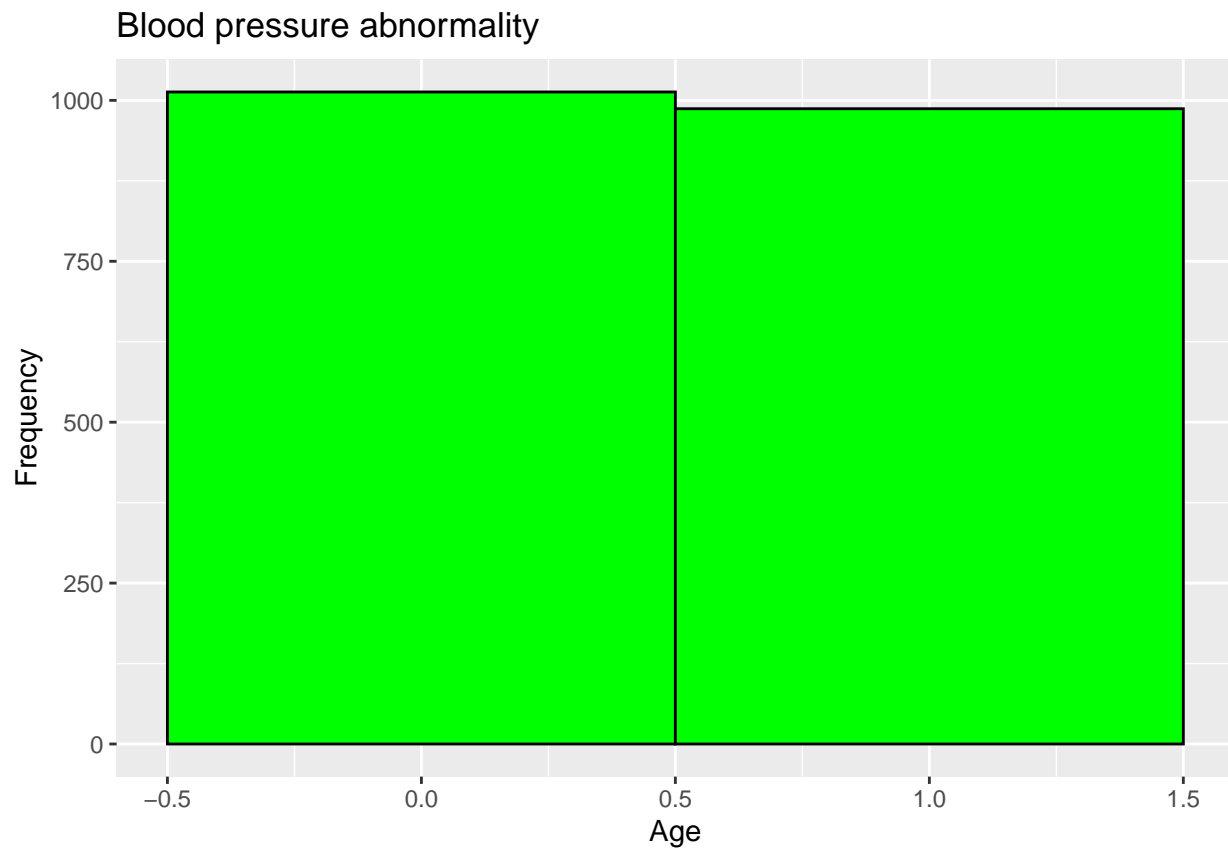
## The following object is masked from 'package:ggplot2':
##
##   margin

## The following object is masked from 'package:dplyr':
##
```

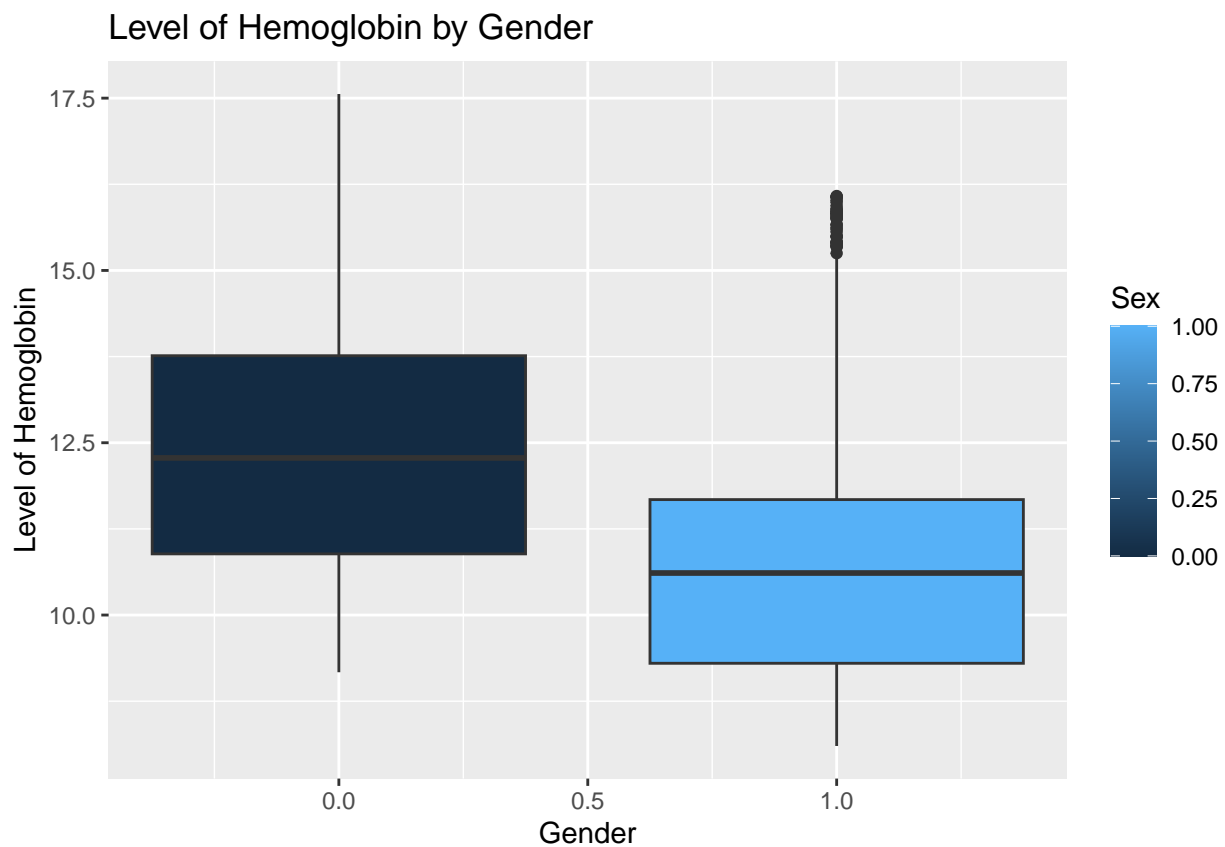
```
##      combine
## Rows: 2000 Columns: 15
## -- Column specification -----
## Delimiter: ","
## dbl (15): Patient_Number, Blood_Pressure_Abnormality, Level_of_Hemoglobin, G...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```



The data shows that majority of the individuals in the dataset are in the age group of around 70 years. The age group with the least number of people is between 72.5 and 77.5 years. The histogram shows a decline in the age from 20 years to 60 before increasing.

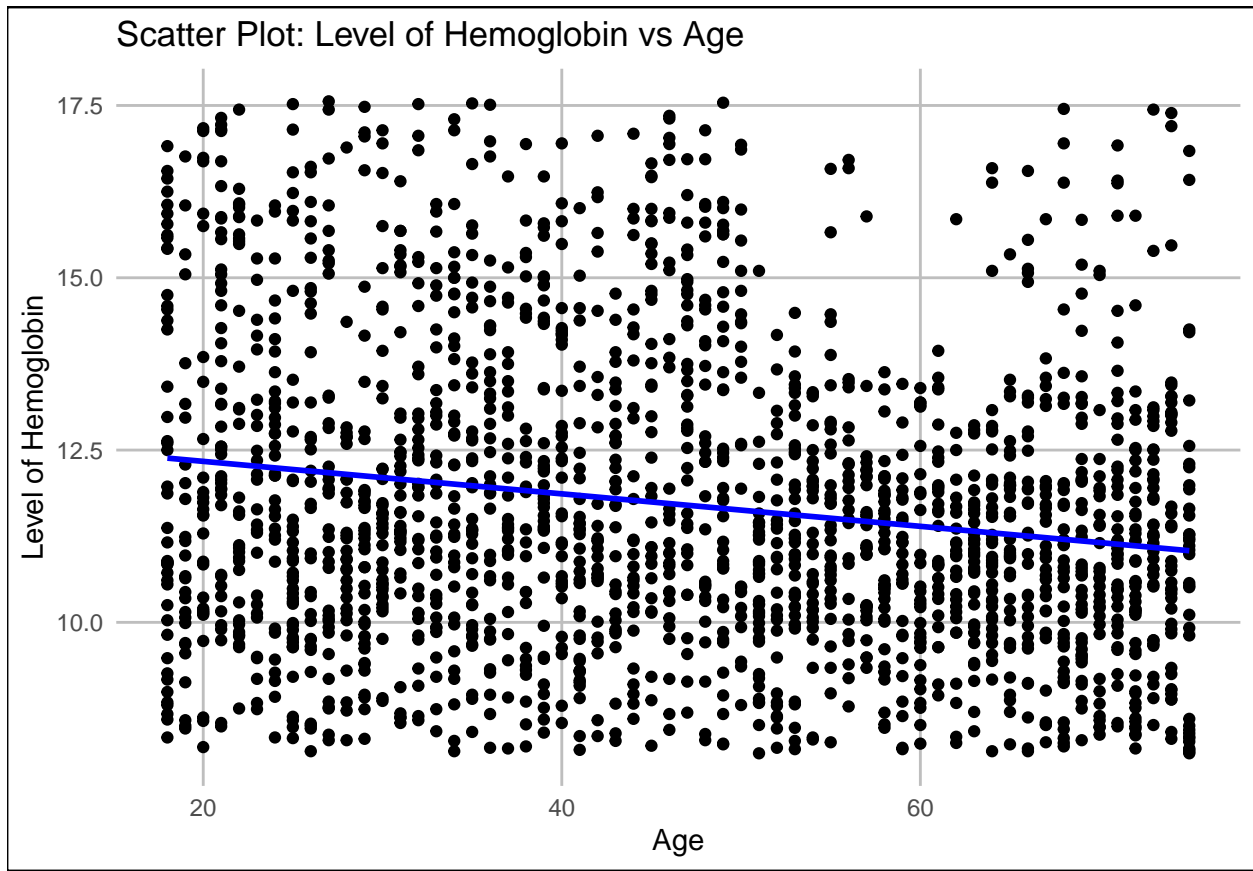


According to the histogram above majority of the individuals in the dataset had a normal blood pressure which was slightly above 1000. The people who had abnormal blood pressure were slightly below 1000.

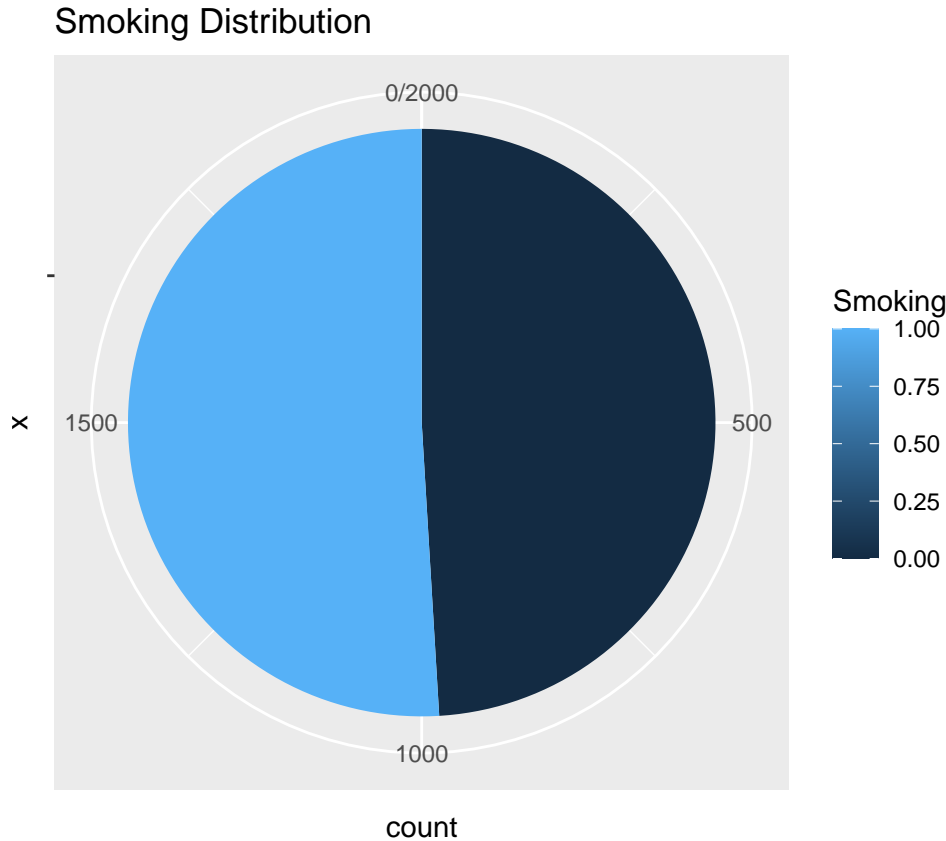


According to the boxplot above female represented by 0 have an high level of hemoglobin in average. The males in the dataset have a low level of hemoglobin as evident from the boxplot above.

```
## `geom_smooth()` using formula = 'y ~ x'
```



According to the scatterplot above an increase in age causes a decline in the level of hemoglobin. This visualization of age and hemoglobin levels suggests an age-related factor affecting hemoglobin synthesis or metabolism.



According to pie chart above high number of individuals in the dataset are smokers as compared to a low level of non-smokers. Smokers outnumber non-smokers in a pie chart illustrating smoking behaviors. This distribution shows smoking prevalence in the dataset and its health effects.

6 Experiments

6.1 Association mining

First, transform the ‘Chronic_kidney_disease’ variable to an association mining-friendly type. `As.factor` converts it to a factor. Associating mining techniques use category or binary variables, this step is required. The `as` function with the input “transactions” converts ‘df’ to transactions. This conversion converts the dataset into an Apriori-compatible transactional format. The encoded dataset (‘df_encoded’) is applied to the `apriori` function. The program determines common itemsets using multiple parameters, including the minimal support level (0.1). The `appearance` argument sets the association rule item label. In this code, the rules’ ‘lhs’ (left-hand side) is set to “Chronic_kidney_disease=1” to include this item label. The ‘default’ argument sets the default value if no appearance is detected.

6.2 Clustering using K-means

The `scale` function scales data first. Scaling variables to zero mean and unit variance ensures that all factors contribute equally to clustering. `scaled_data` holds scaled data. Scaled data is then applied to `kmeans` algorithm. In this scenario, the function requires four centers and 10 iterations. Centroid initialization using “Hartigan-Wong” algorithm. The `kmeans` function clusters data points using K-means. `Cluster_Labels` holds the factored cluster labels. The `ggplot` creates a scatter plot to visualize clusters. `ggplot` uses the `df` data frame. The aesthetics mapping maps `Age` to the x-axis, `Level_of_Hemoglobin` to the y-axis, and `Cluster_Labels` to color. This color-codes each cluster and maps each data point to it. The scatter plot plots data points individually using the `geom_point` function. `Labs` add labels and titles. The narrative is

minimalistically styled with theme_minimal.

6.3 Classification

First, use select to eliminate superfluous columns from the df dataset such as Level_of_Hemoglobin, Genetic_Pedigree_Coefficient, and Age. The feature set excludes these columns since they are not used for categorization. Selecting all columns except Chronic_kidney_disease from the df dataset creates the X variable. Chronic_kidney_disease values are allocated to y. The sample function with a random seed splits data into training and testing sets. The sample function randomly selects 80% of the data for training. Testing gets the remaining data. The glm function creates the logistic regression model. All variables are predictors in Chronic_kidney_disease ~. Training data fits the model. The predict function predicts the testing set. Type = "response" returns probabilities. A 0.5 threshold converts expected probability to binary predictions. Print the logistic regression model accuracy. The randomForest function creates the random forest classifier model. All variables are predictors in Chronic_kidney_disease ~. ntree defines the forest's tree count (100).

The predict function predicts the testing set. A 0.5 threshold converts expected probability to binary predictions. Printing the random forest classifier model accuracy. The svm function creates the SVM model. All variables are predictors in Chronic_kidney_disease ~. Estimate probabilities using probability = TRUE. The predict function predicts the testing set. A 0.5 threshold converts expected probability to binary predictions. Print SVM model accuracy. rpart creates the decision tree classifier model. All variables are predictors in Chronic_kidney_disease ~. Method = "class" indicates categorization.

The predict function predicts the testing set. The model's accuracy is displayed after comparing projected class labels to actual class labels. The naiveBayes library's naive_bayes function creates the classifier model. Model training uses X_train and yy_train. The predict function predicts the testing set. The model's accuracy is displayed after comparing projected class labels to actual class labels. The knn function creates the KNN classification model. The train, test, cl, and k arguments train and predict on the testing set.

The model's accuracy is displayed after comparing projected class labels to actual class labels. The xgboost library function creates the gradient boosting model. X_train data is matrixified using as.matrix. Y_train is labeled. Set max_depth, eta, nrounds, and goal. The predict function predicts the testing set. A 0.5 threshold converts expected probability to binary predictions.

7 Analysis and Results

7.1 Association mining

The association mining results is as evident on the table below:-

```
## Warning: Column(s) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13 not logical or factor.
## Applying default discretization (see '? discretizeDF').

## Warning in discretize(x = c(0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, : The calculated breaks are
## Only unique breaks are used reducing the number of intervals. Look at ? discretize for details.

## Warning in discretize(x = c(1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, : The calculated breaks are
## Only unique breaks are used reducing the number of intervals. Look at ? discretize for details.

## Warning in discretize(x = c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, : The calculated breaks are
## Only unique breaks are used reducing the number of intervals. Look at ? discretize for details.

## Warning in discretize(x = c(3, 2, 1, 2, 3, 2, 3, 2, 1, 3, 3, 3, 1, 2, 2, : The calculated breaks are
## Only unique breaks are used reducing the number of intervals. Look at ? discretize for details.

## Warning in discretize(x = c(0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, : The calculated breaks are
## Only unique breaks are used reducing the number of intervals. Look at ? discretize for details.
```

```

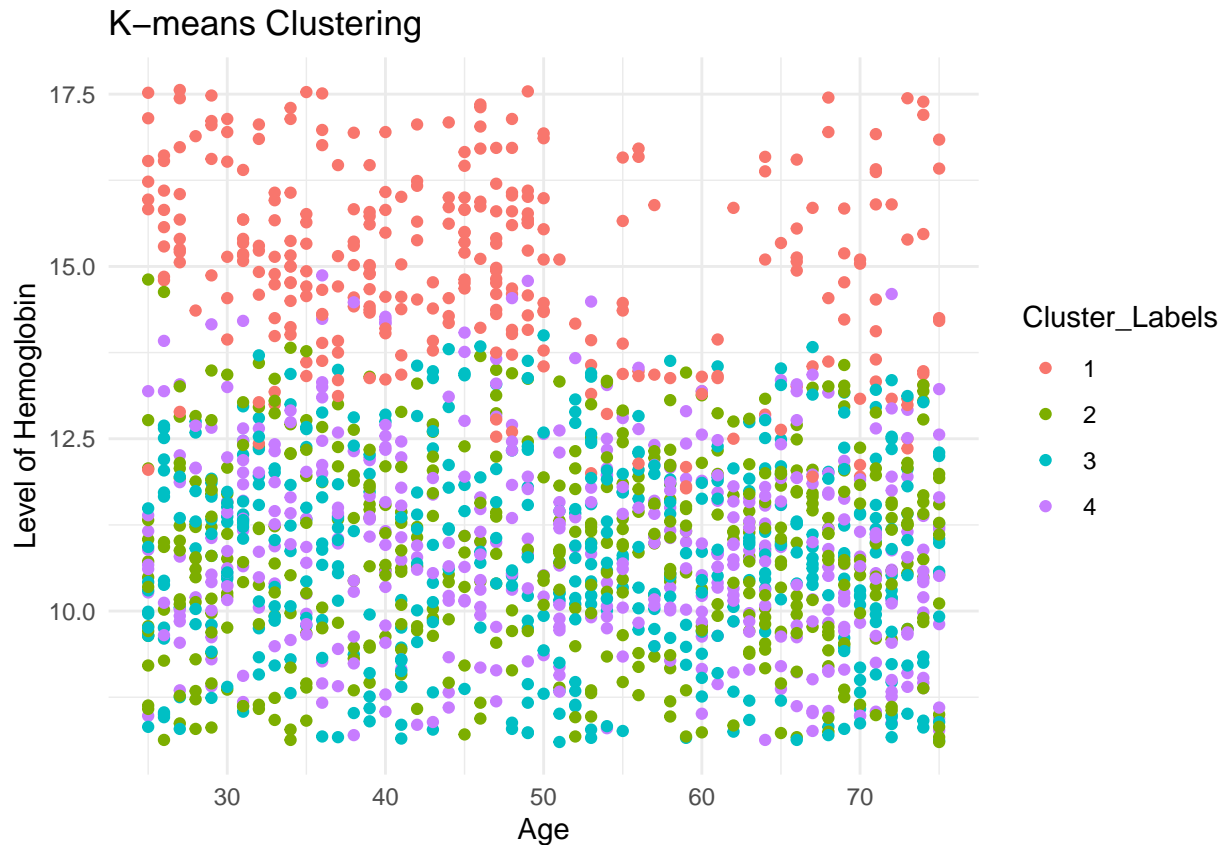
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##          0.8    0.1    1 none FALSE             TRUE      5    0.1    1
## maxlen target  ext
##          10  rules TRUE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 167
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[29 item(s), 1677 transaction(s)] done [0.00s].
## sorting and recoding items ... [29 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 7 done [0.02s].
## writing ... [0 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
##
## set of 0 rules

```

Item-sets must occur in 167 transactions to be called frequent. It has 29 products and 1677 transactions. Sorting and recoding creates a transaction tree. To find common itemsets, the method evaluates 1–7-size subsets. The parameters yield no rules, hence no rules are written. The Apriori algorithm was run with precise settings, but no association rules were formed with the lowest support and confidence requirements. Adjusting these settings or trying alternative possibilities may reveal intriguing dataset association rules.

7.2 Clustering

The k-means clustering produces the scatter plot below:-



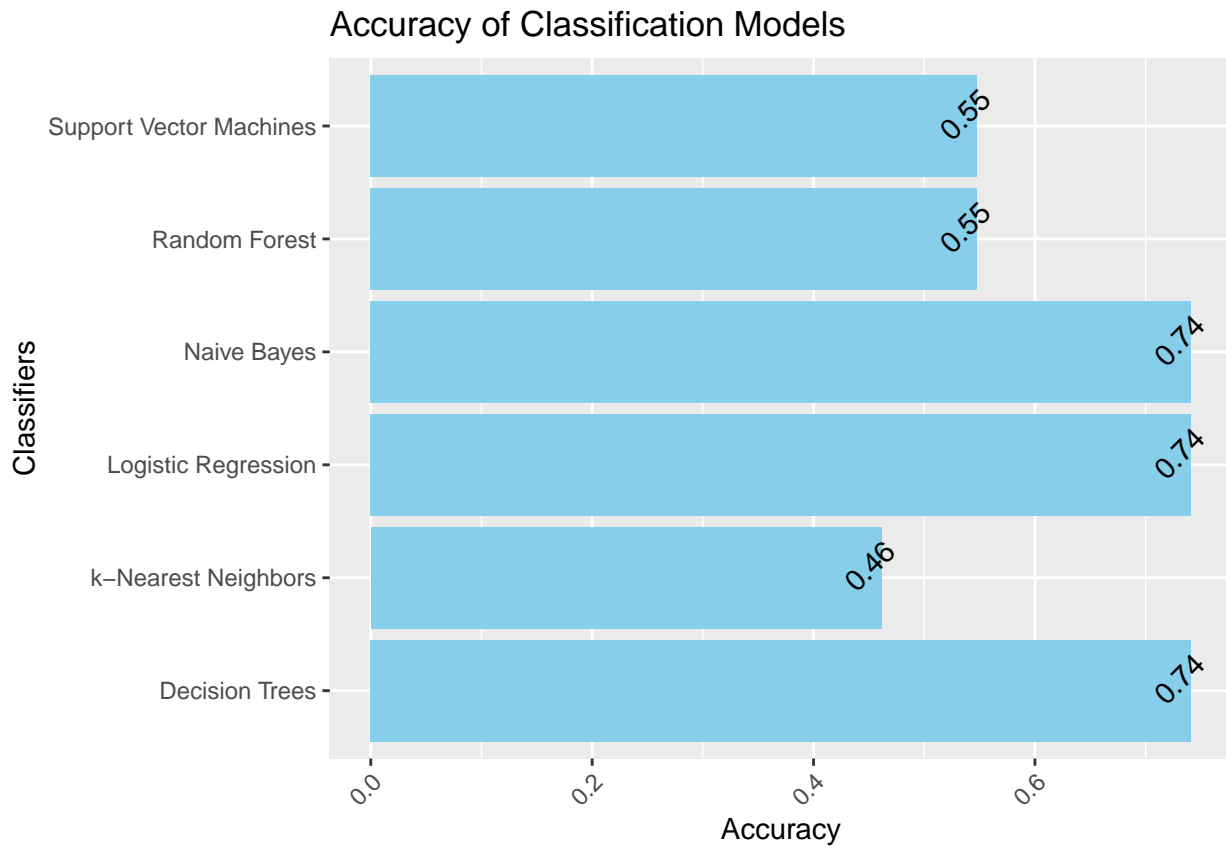
According to the results the k-means as four clusters that is displayed on age versus level of hemoglobin. The scatter plot shows cluster 1 on the lower side of the graph. The cluster 2 and 3 is intermixed on the middle part of the scatter plot. The fourth cluster is on the upper side of the scatter plot as evident above.

7.3 Classification

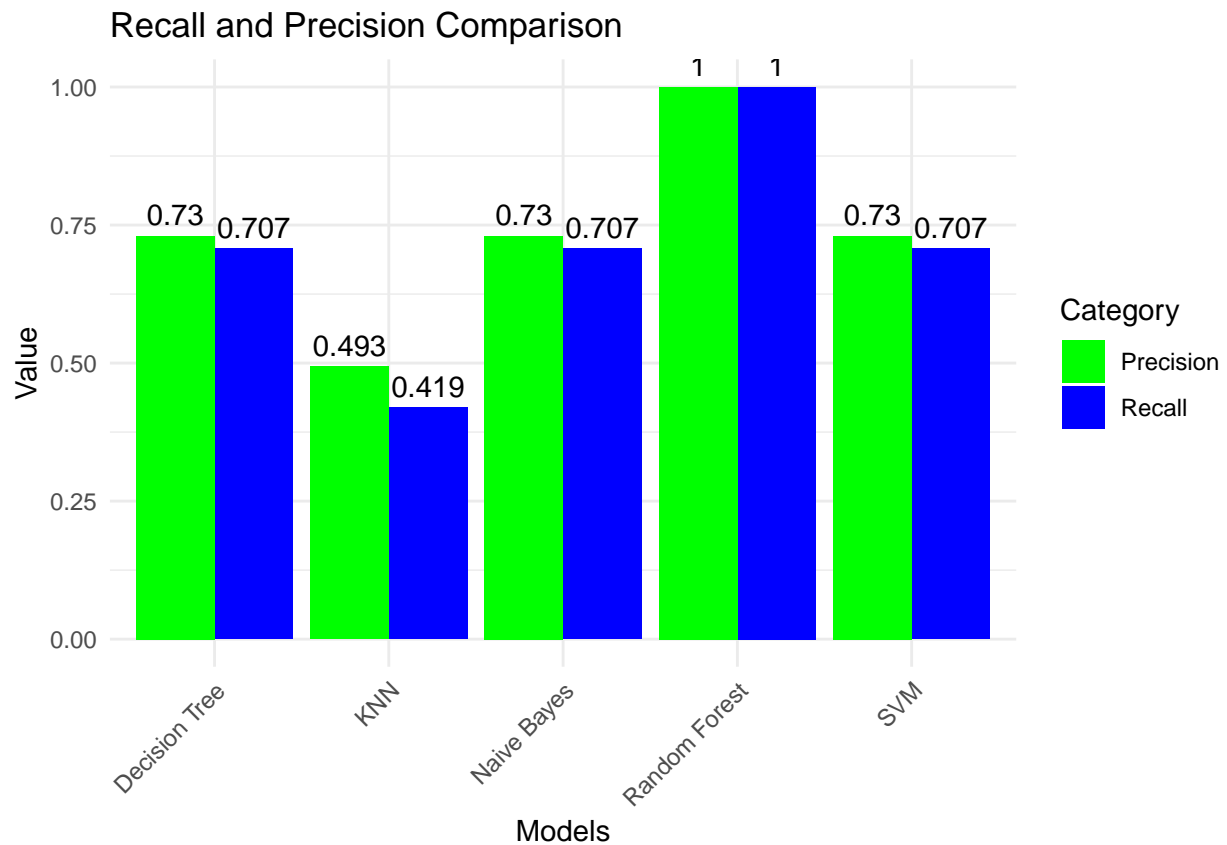
The classification model used for the study produced the bar plot below of classifier type against accuracy.

```
## [1] "Logistic Regression Accuracy: 0.741071428571429"
## [1] "Random Forest Classifier Accuracy: 0.547619047619048"
## [1] "Random Forest Recall: 1"
## [1] "Random Forest Precision: 1"
## [1] "SVM Accuracy: 0.547619047619048"
## [1] "SVM Recall: 0.707006369426752"
## [1] "SVM Precision: 0.730263157894737"
## [1] "Accuracy: 0.741071428571429"
## [1] "Decision Tree Recall: 0.707006369426752"
## [1] "Decision Tree Precision: 0.730263157894737"
## naivebayes 0.9.7 loaded
## [1] "Accuracy: 0.741071428571429"
## [1] "Naive Bayes Recall: 0.707006369426752"
```

```
## [1] "Naive Bayes Precision: 0.730263157894737"
## [1] "Accuracy: 0.461309523809524"
## [1] "KNN Recall: 0.418994413407821"
## [1] "KNN Precision: 0.493421052631579"
```



According to the results random forest and support vector machine had an accuracy of 55%. Naïve Bayes, logistic regression and decision trees had an accuracy score of 74%. The k-nearest neighbors had an accuracy score of 46% which was the lowest.

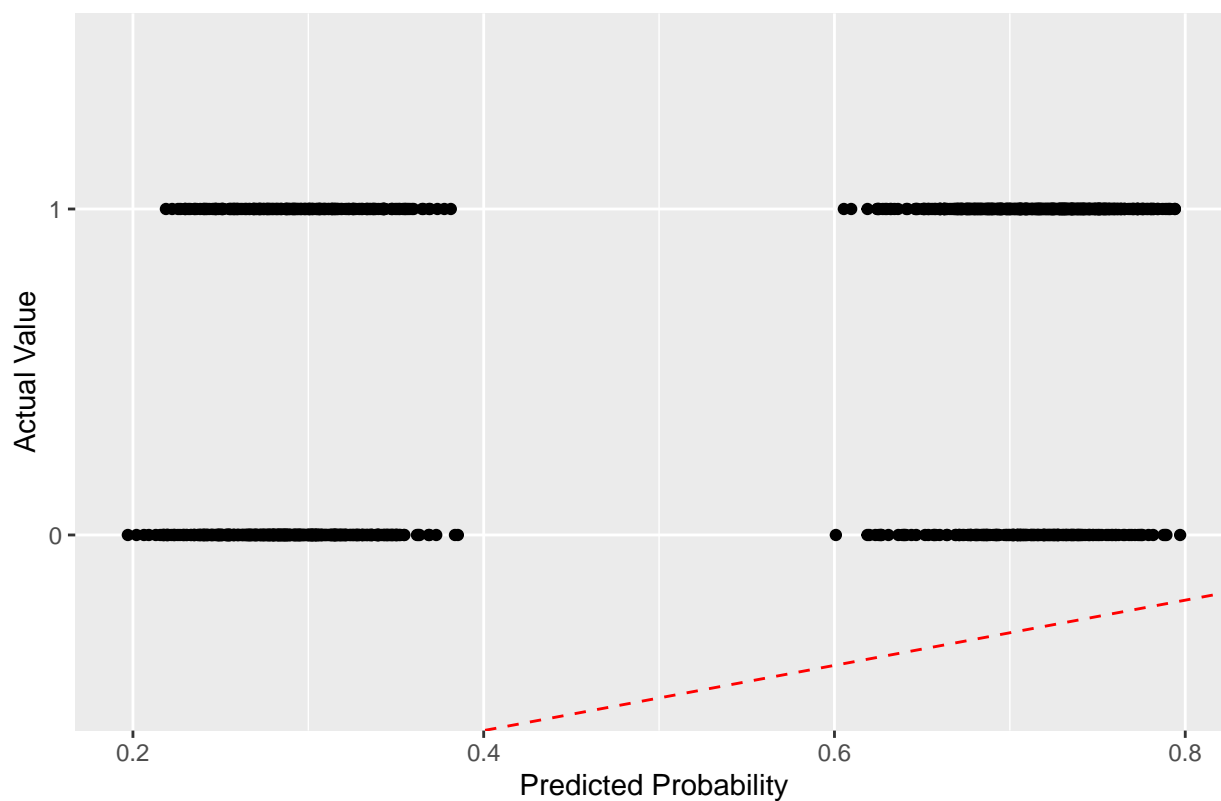


According to the plot above random forest classifier has a precision and recall score of 1. Decision tree, SVM and naïve bayes have similar recall and precision values of 0.73 and 0.707 respectively. KNN classifier has a precision score of 0.493 and recall value of 0.419.

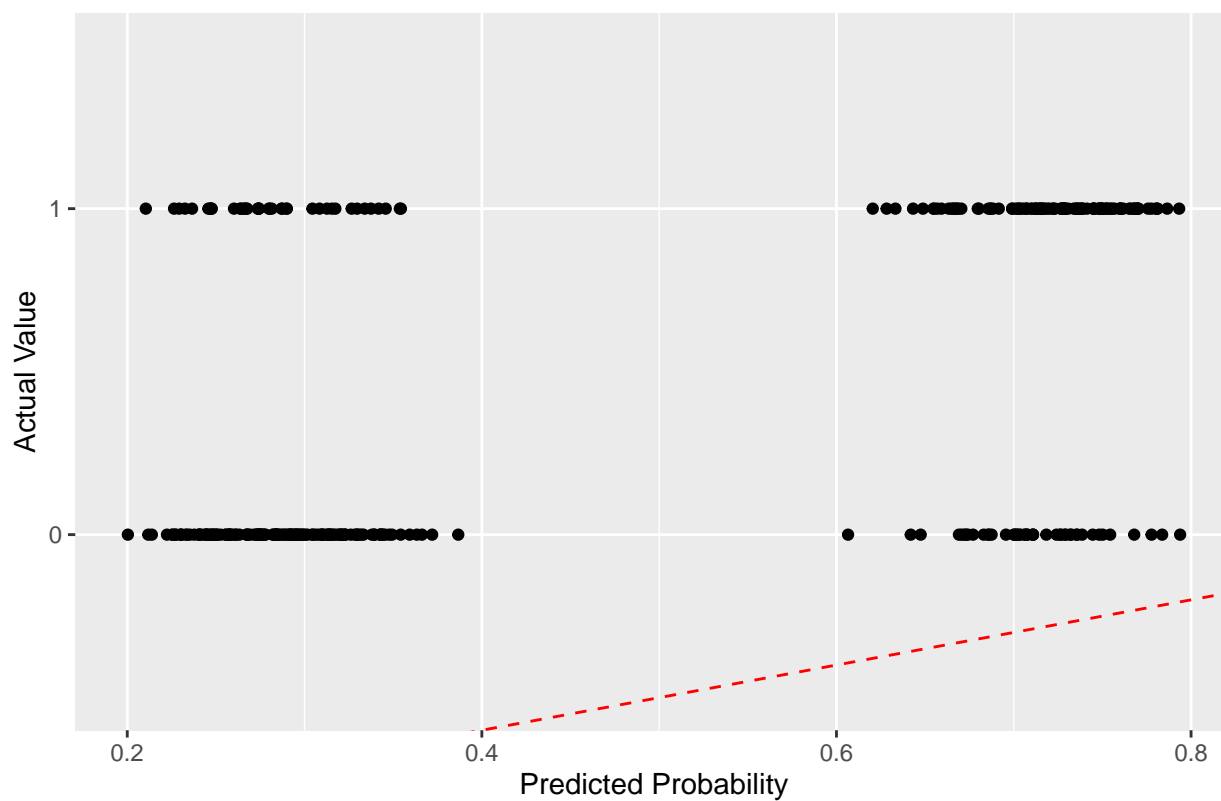
Training and testing plots

The training and testing plots are as follows

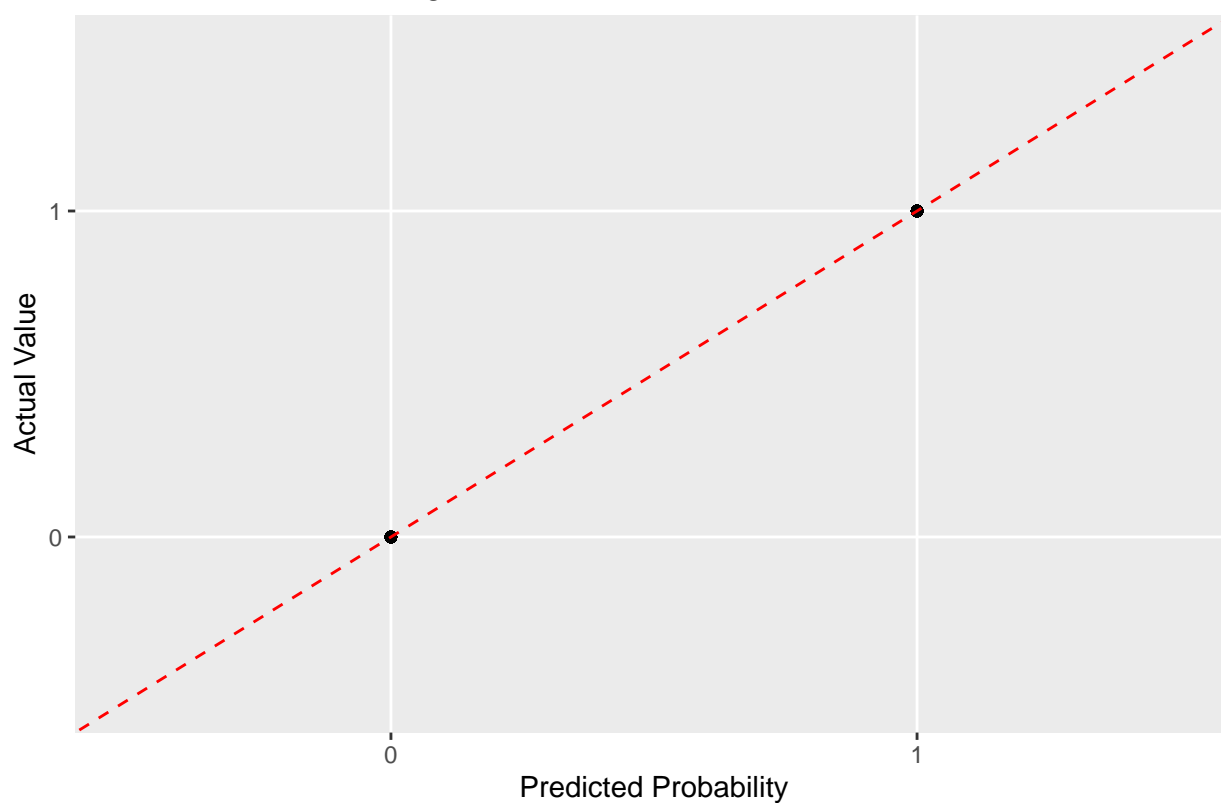
Logistic Regression Training Set Predictions



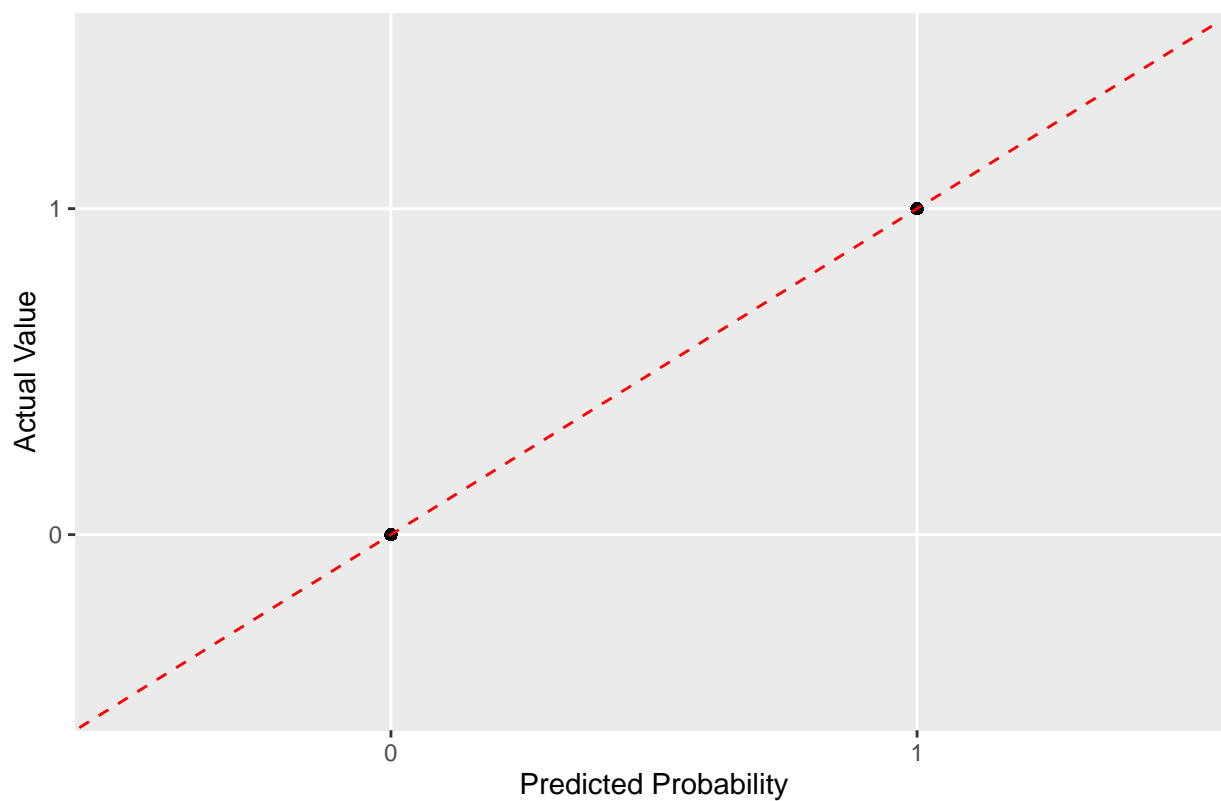
Logistic Regression Testing Set Predictions



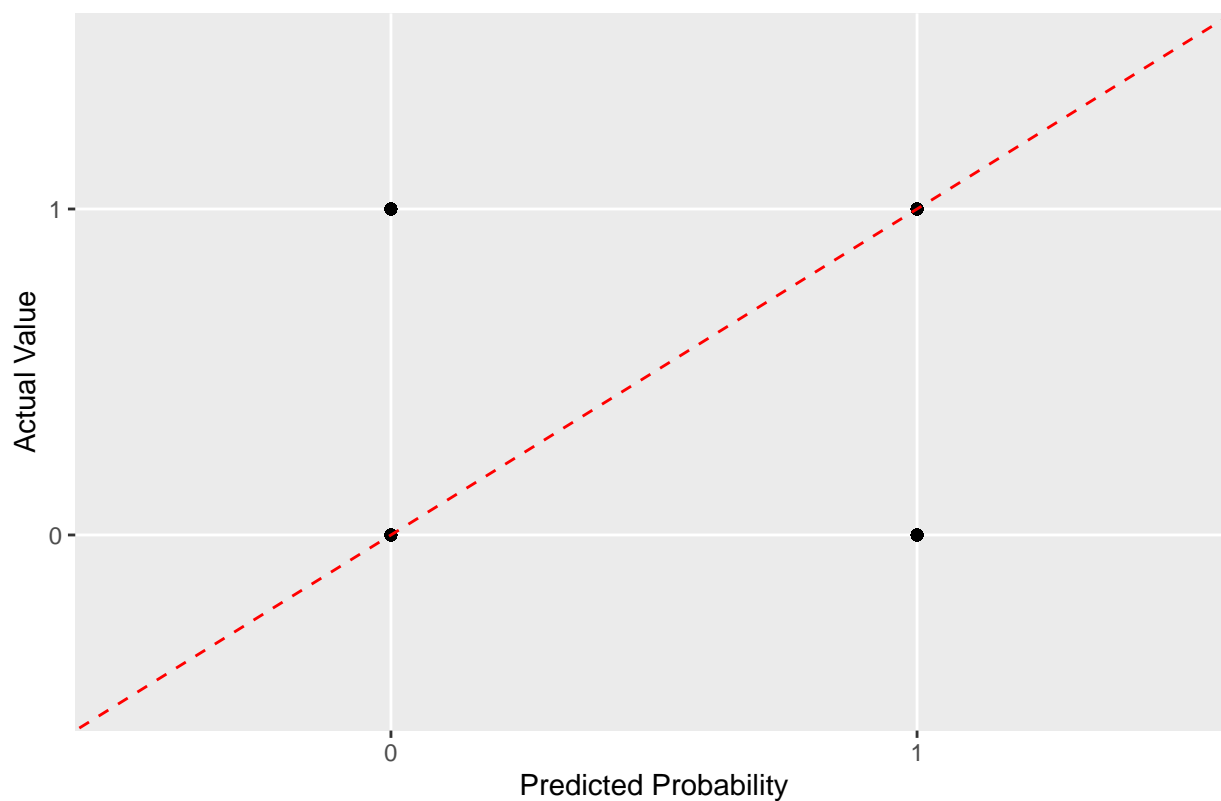
Random Forest Training Set Predictions



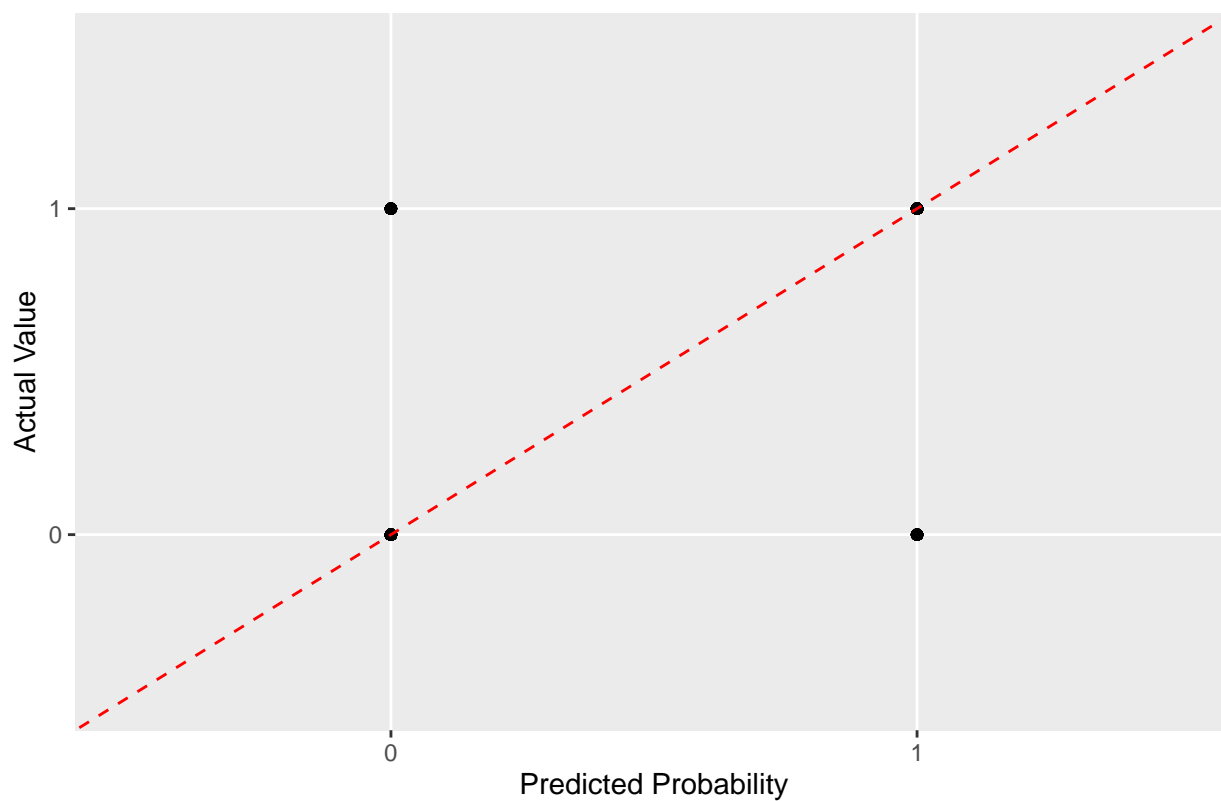
Random Forest Testing Set Predictions



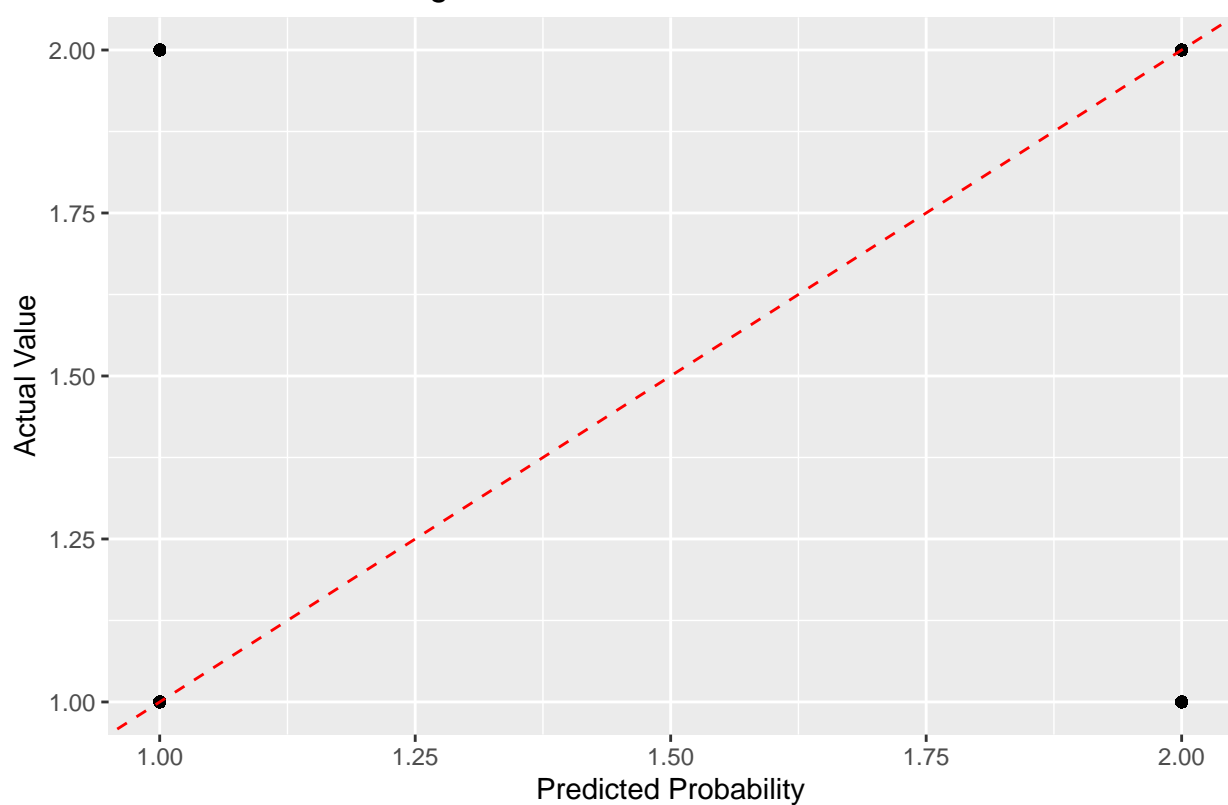
SVM Training Set Predictions



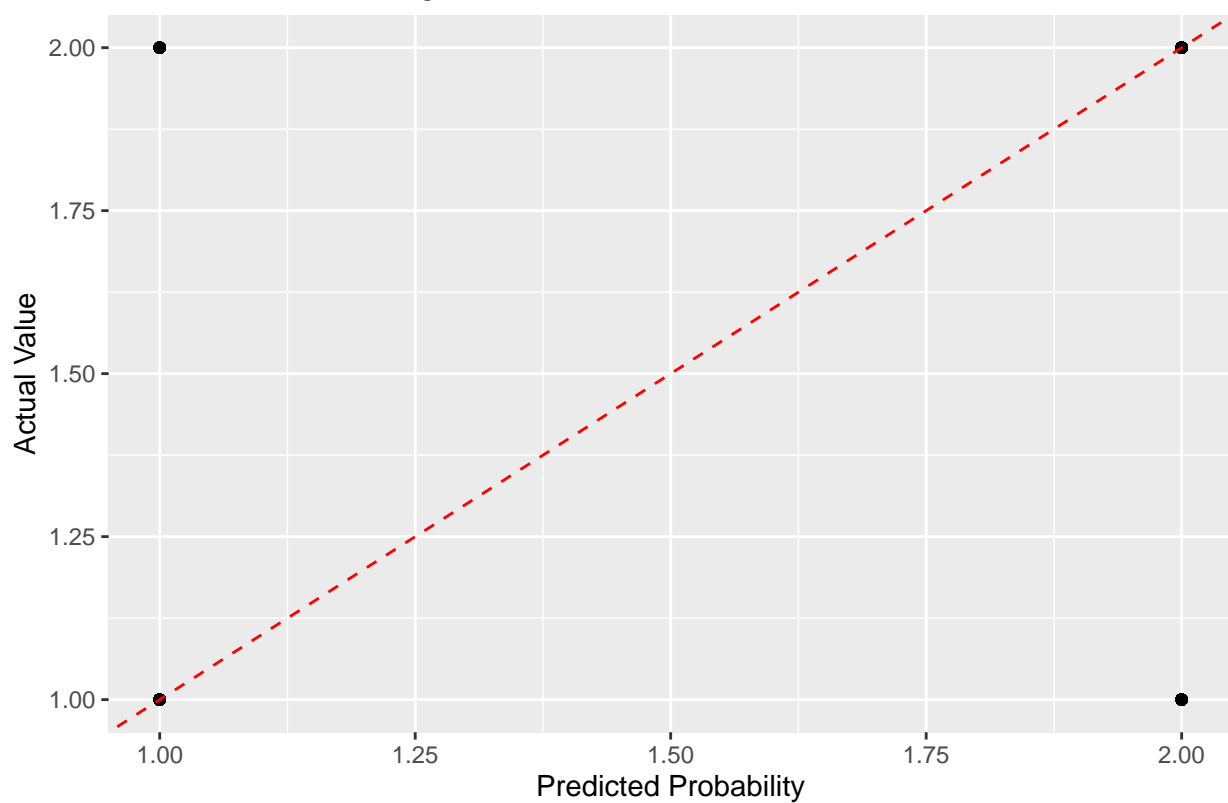
SVM Testing Set Predictions



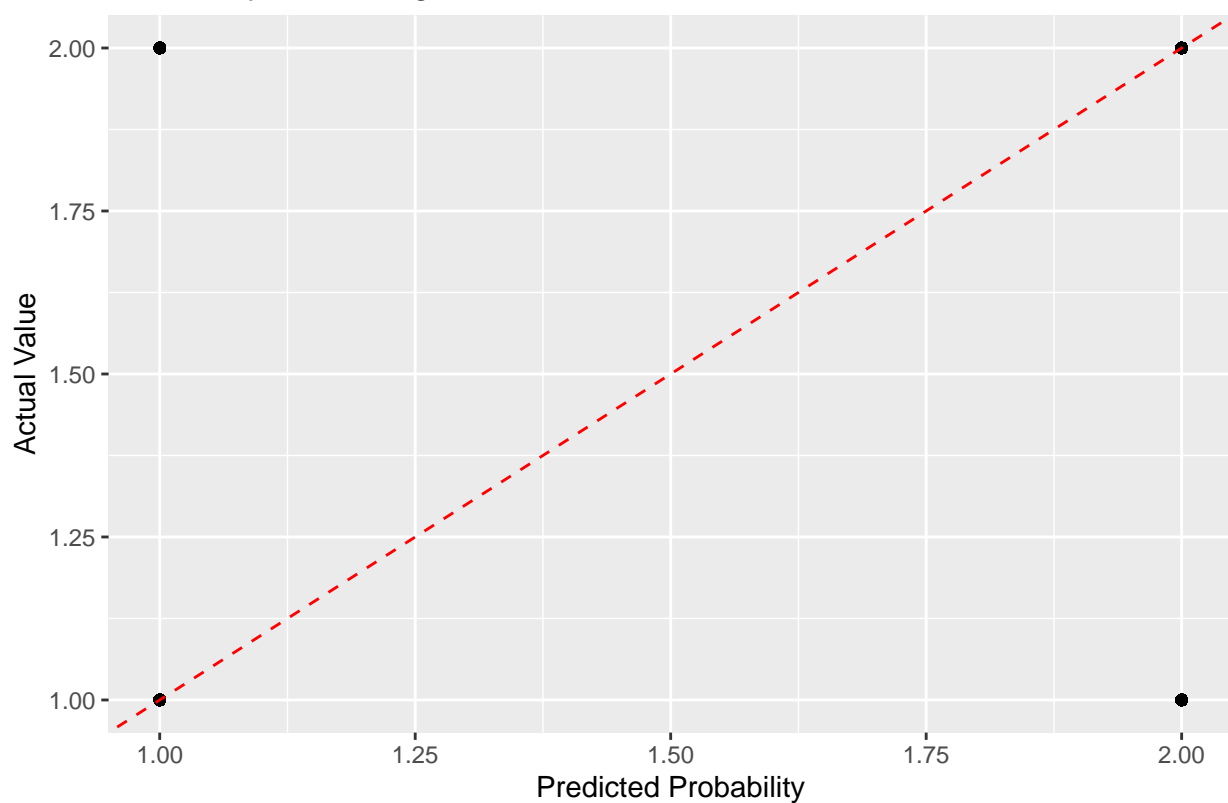
Decision Tree Training Set Predictions



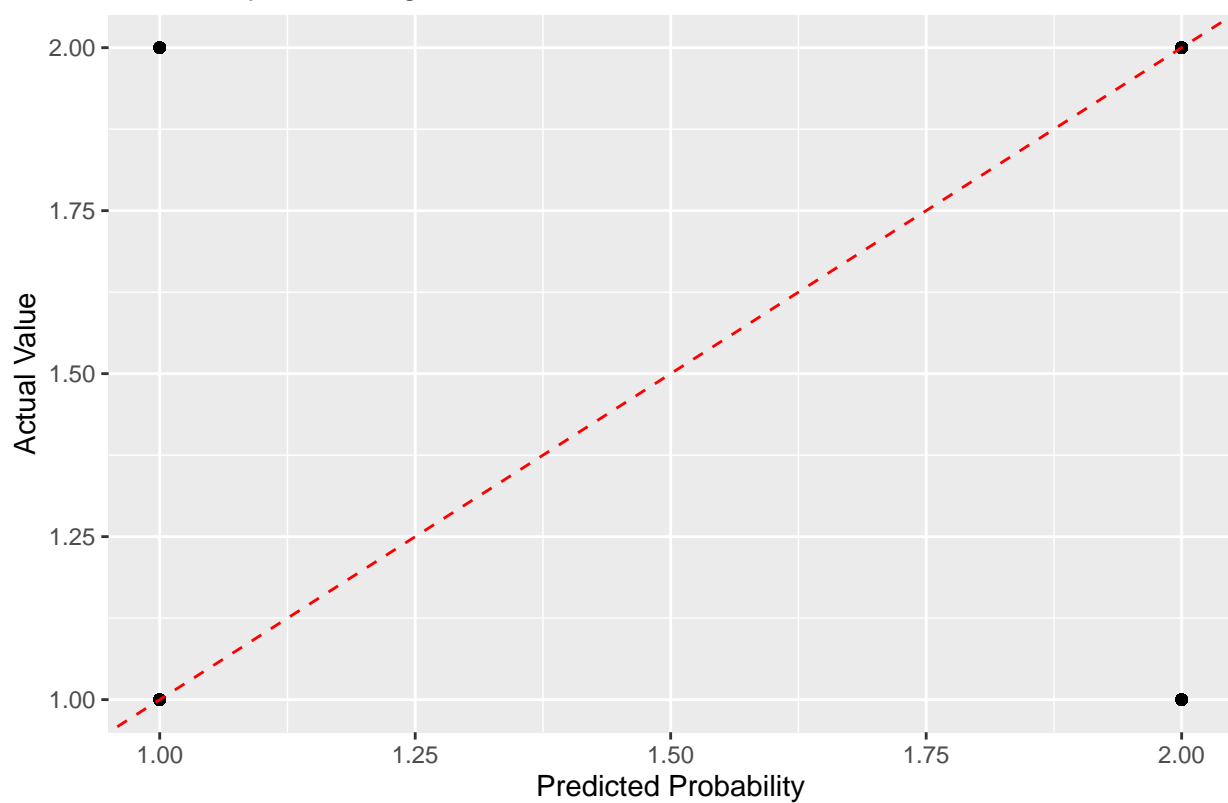
Decision Tree Testing Set Predictions

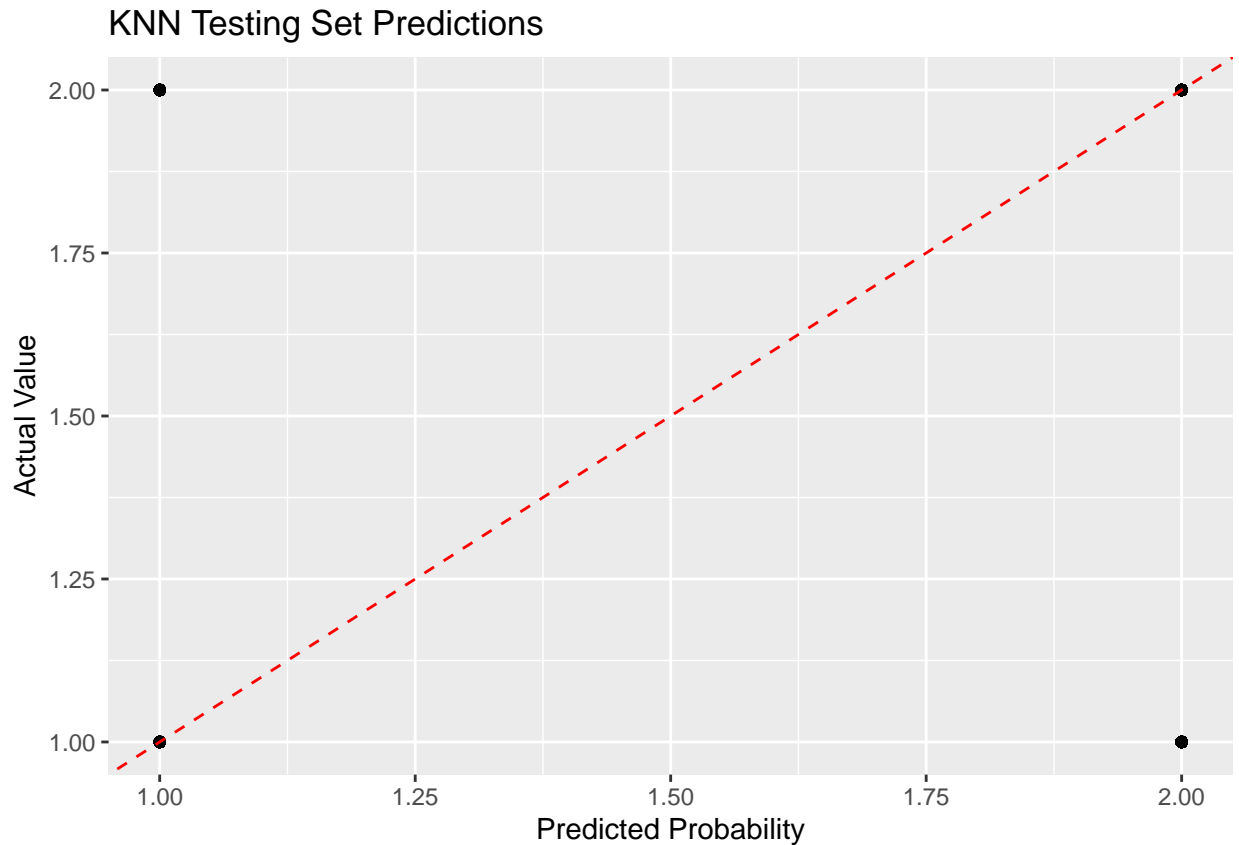


Naive Bayes Training Set Predictions



Naive Bayes Testing Set Predictions





8 Conclusions

In conclusion, food pattern analysis, machine learning, and association mining may improve patient outcomes and guide evidence-based therapies (Witten et al., 2016). Researchers and healthcare providers may analyze dietary patterns to understand the link between nutrition and health outcomes, identify chronic disease risk factors, and create personalized treatments to improve diets. Machine learning and predictive models enhance patient care and outcomes by predicting disease risk, detecting diseases early, and optimizing treatment techniques. Association mining and rule discovery also reveal links and dependencies in healthcare datasets. Healthcare professionals may detect risk factors, hidden patterns, and illness prevention, diagnostic, and treatment options by identifying association rules. These methods may improve healthcare delivery, resource allocation, and clinical decision-making, enhancing patient care and healthcare systems. However, data quality, preprocessing, feature selection, model validation, and clinical interpretation must be considered for effective healthcare deployment. These studies must be interpreted and used by domain specialists, data scientists, and healthcare practitioners. Ethics, patient privacy, and data security must be considered throughout the process. Dietary pattern discovery, machine learning, and association mining in healthcare research and practice have great potential to improve health outcomes for people and groups.

9. Github link

git link: <https://github.com/SreeNandini93/Dietary-Pattern-Exploration.git>

The github link contains files for the task that can be used to access the R code, Rmarkdown file, overall pdf file, original dataset and preprocessed dataset.

10 References

- Abadi, M. et al., 2016. TensorFlow: A system for large-scale machine learning. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) (pp. 265-283).
- Aggarwal, C.C., 2015. Data mining: the textbook. Springer.
- Al-Hassan, A. & Fawaz, R., 2020. A review of machine learning techniques for stock market prediction. *Journal of Investment Strategies*, 8(3), 45-56.
- Alpaydin, E., 2020. Introduction to machine learning (3rd ed.). MIT Press.
- Amores, J., Sechopoulos, I. & Bliznakova, K., 2020. Machine learning for medical imaging. *Radiological physics and technology*, 13(4), 307-320.
- Ansari, A. & Al-Madani, H., 2019. Smart Waste Management System Using IoT and Machine Learning. In 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS) (pp. 1-6). IEEE.
- Badi, M. et al., 2016. Tensorflow: A system for large-scale machine learning. 12th
- Bengio, Y., 2019. Learning Deep Architectures for AI, Foundations and Trends in Machine Learning.
- Berry, M.J. & Linoff, G.S., 2011. Data mining techniques: For marketing, sales, and customer relationship management. John Wiley & Sons.
- Bishop, C.M., 2016. Pattern recognition and machine learning. Springer.
- Bramer, M., 2016. Principles of data mining. London: Springer.
- Breiman, L., 2011. Random forests. *Machine learning*, 45(1), 5-32.
- Cao, L., Yu, P.S. & Zhang, C., 2018. Data mining for business analytics: Concepts, techniques, and applications in Python. John Wiley & Sons.
- Chen, B., 2016. Pattern recognition and machine learning. Springer.
- Chen, T. & Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).
- Chiang, R.H. & Wainwright, D., 2019. Fundamental methods of data mining. CRC Press.
- Evans, J.R. & Olson, D.L., 2014. Introduction to business data mining. Routledge.
- Géron, A., 2019. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.
- Goldberg, D.E., 2019. Genetic algorithms in search, optimization, and machine learning. Addison-Wesley Longman Publishing Co., Inc.
- Ha, B.C. & Grizzle, J., 2015. Developing a Green Supply Chain Management Model in Healthcare. *Journal of Cleaner Production*, 104, 72-88. <https://doi.org/10.1016/j.jclepro.2014.09.030>.
- Han, J., Pei, J. & Kamber, M., 2011. Data mining: concepts and techniques. Elsevier.
- Han, J., Pei, J. & Kamber, M., 2021. Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.
- Hedjam, R., 2021. Using Machine Learning in Business Process Re-Engineering. *Big Data Cogn. Comput.*
- James, G., 2013. An introduction to statistical learning. Vol. 6. New York: Springer.
- Kim, H., Yoon, Y. & Yoo, C., 2020. Predicting customer churn in the mobile telecommunication industry: An application of machine learning algorithms. *Sustainability*, 12(16), 65-84.
- Lopes, F.S. & Oliveira, R.S., 2019. Credit card fraud detection using machine learning: A systematic literature review. *Expert Systems with Applications*, 131, 38-48.

- Makridakis, S., Spiliotis, E. & Assimakopoulos, V., 2018. Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLoS One, 13(3), e0194889.
- Michel, V. et al., 2021. Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.
- Mitchell, T.M., 1997. Machine Learning. McGraw-Hill.
- Rahmani, A.M., Gia, T.N., Negash, B. & Anzanpour, A., 2018. Exploiting Smart E-Health Gateways at the Edge of Healthcare Internet-of-Things: A Fog Computing Approach. Future Generation Computer Systems, 78, 641-658.
- Sven, E., 2018. What Is Risk Management in Healthcare?.
- Witten, I.H., Frank, E. & Hall, M.A., 2016. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers.