**HEADER** Who:KK What: Naive base algoritham Last edited: 2023-04-08 **CONTENTS** Introduction Naive bayes setup Introduction Precision agriculture is in trend nowadays. It helps the farmers to get informed decision about the farming strategy. The science of training machines to learn and produce models for future predictions is widely used, and not for nothing. Agriculture plays a critical role in the global economy. With the continuing expansion of the human population understanding worldwide crop yield is central addressing food security challenges and reducing the impacts of climate change. Crop yield prediction is an important agricultural problem. The Agricultural yield primarily depends on weather conditions (rain, temperature, etc), pesticides and accurate information about history of crop yield is an important thing for making decisions related to agricultural risk management and future predictions. N - ratio of Nitrogen content in soil P - ratio of Phosphorous content in soil K - ratio of Potassium content in soil temperature - temperature in degree Celsius humidity - relative humidity in % ph - ph value of the soil rainfall - rainfall in mm label - type of crop Dataset source: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset Naive bayes Naive Bayes is a Supervised Machine Learning algorithm based on the Bayes Theorem that is used to solve classification problems by following a probabilistic approach. It is based on the idea that the predictor variables in a Machine Learning model are independent of each other. Meaning that the outcome of a model depends on a set of independent variables that have nothing to do with each other. Problem statement Probabilities of which type of crops is suitable for cultivation based on the criteria such as ph, rainfall, temperature, etc... Setup setwd("C:/Users/DELL/Downloads") getwd() ## [1] "C:/Users/DELL/Downloads" set working directory to import our dataset crop <- read.csv("Crop\_recommendation.csv")</pre> Read the data into R environment Checking Null values library(Amelia) ## Warning: package 'Amelia' was built under R version 4.2.3 ## Loading required package: Rcpp ## ## ## ## Amelia II: Multiple Imputation ## ## (Version 1.8.1, built: 2022-11-18) ## ## Copyright (C) 2005-2023 James Honaker, Gary King and Matthew Blackwell ## ## Refer to http://gking.harvard.edu/amelia/ for more information ## ## missmap(crop) Missingness Map 2170 2080 1990 1900 1810 1720 1630 1540 1450 1360 1270 1180 1090 1000 Missing (0%) 910 Observed (100%) 820 730 640 550 460 370 280 190 Explore the data summary(crop) Ν K temperature Min. : 0.00 Min. : 5.00 Min. : 5.00 Min. : 8.826 1st Qu.:22.769 Median : 37.00 Median : 51.00 Median : 32.00 Median :25.599 Mean : 50.55 Mean : 53.36 Mean : 48.15 Mean : 25.616 ## 3rd Qu.: 84.25 3rd Qu.: 68.00 3rd Qu.: 49.00 3rd Qu.:28.562 Max. :140.00 Max. :145.00 Max. :205.00 Max. :43.675 ## humidity ph rainfall ## label :14.26 Min. :3.505 Min. : 20.21 Length:2200 Median :80.47 Median :6.425 Median : 94.87 Mode :character Mean :71.48 Mean :6.469 Mean :103.46 3rd Qu.:89.95 3rd Qu.:6.924 3rd Qu.:124.27 Max. :99.98 Max. :9.935 Max. :298.56 names(crop) ## [1] "N" "temperature" "humidity" ## [6] "ph" "rainfall" Checking the correlation for pairwise among all the predicators in the dataset. cor(crop[ , -8]) K temperature humidity 1.00000000 -0.23145958 -0.14051184 0.02650380 0.190688379 -0.23145958 1.00000000 0.73623222 -0.12754113 -0.118734116 ## N ## P ## K -0.14051184 0.73623222 1.00000000 -0.16038713 0.190858861 ## temperature 0.02650380 -0.12754113 -0.16038713 1.00000000 0.205319677 ## temperature -0.017795017 -0.03008378 ## humidity -0.008482539 0.09442305 ## ph 1.000000000 -0.10906948 ## rainfall -0.109069484 1.00000000 library(e1071) #for naive bayes algoritham ## Warning: package 'e1071' was built under R version 4.2.3 library(caret) #for data preprocessing ## Warning: package 'caret' was built under R version 4.2.3 ## Loading required package: ggplot2 ## Warning: package 'ggplot2' was built under R version 4.2.3 ## Loading required package: lattice library(caTools) # split the data for train and test ## Warning: package 'caTools' was built under R version 4.2.3 split the data for test and train set.seed(1) split <- sample.split(crop\$label, SplitRatio = 0.7)</pre> train\_data <- subset(crop, split == "TRUE")</pre> test\_data <- subset(crop, split =="FALSE")</pre> Deploy the naive bayes model set.seed(123) # For reproduceability result <- naiveBayes(label ~ ., data = train\_data, family = "multinomial")</pre> ## Naive Bayes Classifier for Discrete Predictors ## ## naiveBayes.default(x = X, y = Y, laplace = laplace, family = "multinomial") ## A-priori probabilities: ## Y ## banana blackgram chickpea coffee apple coconut  $0.04545455 \quad 0.04545455 \quad 0.04545455 \quad 0.04545455 \quad 0.04545455 \quad 0.04545455$ ## cotton jute kidneybeans lentil maize grapes ## 0.04545455 0.04545455 0.04545455 0.04545455 0.04545455 0.04545455mothbeans mungbean muskmelon mango orange papaya 0.04545455 0.04545455 0.04545455 0.04545455 0.04545455 0.04545455 pigeonpeas pomegranate rice watermelon 0.04545455 0.04545455 0.04545455 0.04545455 ## Conditional probabilities: ## ## Y [,1] [,2] 20.47143 11.46099 apple ## banana 101.34286 11.11454 ## blackgram 40.01429 12.97656 39.38571 11.69577 chickpea ## coconut 22.31429 12.39862 ## coffee 100.84286 11.78015 cotton 118.01429 11.09053 23.10000 12.72923 ## grapes ## jute 77.67143 10.35419 kidneybeans 21.11429 11.36170 lentil ## 19.71429 12.22630 maize 78.50000 12.13182 ## mango 18.87143 11.84798 ## mothbeans 19.57143 10.83683 mungbean 21.54286 11.31715 ## muskmelon 100.25714 11.78268 ## orange 19.50000 12.54413 49.08571 12.08753 papaya pigeonpeas 20.71429 11.41718 ## pomegranate 19.24286 12.08477 ## rice 80.32857 12.09590 ## watermelon 100.01429 12.39506 ## ## ## Y [,1] [,2] 133.58571 8.064042 apple 82.22857 7.600861 ## banana 67.55714 7.377278 ## blackgram chickpea 67.78571 7.208877 17.14286 7.947811 ## coconut ## coffee 28.82857 7.211045 ## cotton 46.37143 7.635105 grapes ## 132.58571 7.616711 ## jute 46.41429 6.925050 kidneybeans 66.47143 7.399023 ## lentil 68.60000 7.157443 48.17143 7.938115 maize mango 26.62857 7.061164 mothbeans 48.11429 7.339674 mungbean 47.57143 7.810382 17.88571 7.472739 muskmelon ## orange 16.57143 7.871378 58.90000 6.936879 papaya pigeonpeas 67.64286 7.079335 pomegranate 18.04286 7.911532 48.50000 7.414733 ## rice watermelon 17.00000 7.385846 ## ## ## Y [,1] [,2] 199.914286 3.322051 apple ## banana 50.071429 3.410646 blackgram 19.342857 3.322674 79.985714 3.299068 chickpea 30.585714 3.113751 coconut 30.000000 3.088079 ## coffee cotton 19.314286 3.052881 ## grapes 199.828571 3.370681 ## jute 40.014286 3.511855 kidneybeans 20.157143 3.237622 lentil 19.214286 2.938465 maize 19.771429 2.969270 mango 30.300000 3.085027 ## mothbeans 20.085714 3.119564 19.685714 3.187579 mungbean muskmelon 50.285714 3.341192 ## orange 9.957143 3.168851 49.985714 2.985437 papaya pigeonpeas 20.200000 2.902273 pomegranate 39.985714 3.019228 40.114286 2.891982 ## ## watermelon 50.214286 3.274375 ## ## temperature ## Y [,1] [,2] 22.59058 0.7946282 apple banana 27.28301 1.4643216 blackgram ## 30.04764 2.6667831 chickpea 18.86736 1.1477175 coconut 27.41991 1.3437954 ## coffee 25.63910 1.4969598 cotton 24.00747 1.1001130 ## grapes 23.21029 9.8418654 25.07158 1.1576610 ## jute kidneybeans 20.37034 2.7211267 lentil 24.23728 3.3015803 maize 22.41417 2.6795308 31.10622 2.6618321 mango mothbeans 28.01206 2.1221696 mungbean 28.47549 0.8327194 muskmelon 28.65453 0.8434818 ## orange 23.39035 6.8534942 33.96815 6.2694970 papaya pigeonpeas 28.02477 5.9287863 pomegranate 22.03030 2.2259221 23.66992 1.9915774 ## rice ## watermelon 25.45188 0.8502784 ## ## humidity ## Y [,1] [,2] 92.30501 1.540897 apple

## 80.30384 2.977812 banana 65.20568 2.820002 blackgram chickpea 16.70433 1.738737 95.00263 2.669986 coconut 59.74336 5.661566 coffee cotton 79.82885 3.088613 ## grapes 81.78503 1.147129 79.72386 5.599977 ## jute kidneybeans 21.74186 2.218724 lentil 65.23182 2.930849 64.70105 5.427350 50.19352 2.858317 ## mango ## mothbeans 53.06803 7.214091 mungbean 85.32290 2.850170 muskmelon 92.19570 1.481342 ## orange 91.98953 1.367901 ## papaya 92.36651 1.311087 pigeonpeas 47.72933 10.792016 pomegranate 90.20402 2.767234 82.36622 1.421458 watermelon 85.29873 2.959802 ## ## ## ## Y [,1] [,2] 5.943242 0.2622972 apple 5.983609 0.2751254 ## banana blackgram 7.098010 0.3762074 chickpea 7.221399 0.7620864 5.962484 0.3002321 coconut ## coffee 6.782642 0.3948700 cotton 6.916142 0.6170446 ## grapes 6.032760 0.2971168 ## jute 6.706524 0.4500477 kidneybeans 5.742531 0.1498251 ## lentil 6.958282 0.5321713 ## maize 6.230769 0.4042548 mango 5.748830 0.6874668 ## mothbeans 6.683158 1.8746743 mungbean 6.704140 0.2843508 muskmelon 6.363593 0.2399262 ## orange 6.979031 0.5731536 papaya 6.744258 0.1581820 pigeonpeas 5.776550 0.8716599 ## pomegranate 6.403307 0.4977543 6.312567 0.8023866 watermelon 6.479607 0.2753886 ## ## ## rainfall ## Y [,1] [,2] 113.63155 7.054731 apple 104.94099 9.400104 banana 67.72701 4.268109 blackgram 80.60085 8.059805 chickpea ## coconut 175.09968 29.070461 ## coffee 161.08023 25.796200 cotton 81.26426 11.439471 ## grapes 69.55849 3.117604 ## jute 174.37404 14.559620 kidneybeans 104.30965 25.527117 ## lentil 45.13603 5.978576 ## maize 84.81618 15.552570 ## mango 94.66879 3.288246 ## mothbeans 50.65568 13.903725 mungbean 48.22746 6.723748 muskmelon 24.48847 2.838499 ## orange 110.27951 5.975064 papaya 146.18454 64.691531 pigeonpeas 149.63188 32.741976 pomegranate 107.48176 2.701954 235.86817 34.346133 watermelon 51.22960 5.862716 predict the model predictions <- predict(result, newdata = test\_data)</pre> confusion matrix to check the accuracy of the model cm <- table(test\_data\$label, predictions)</pre> predictions ## apple banana blackgram chickpea coconut coffee cotton grapes jute ## apple 0 banana ## blackgram 0 ## chickpea 0 0 30 0 coconut 0 ## coffee ## 0 0 cotton ## grapes 30 ## 30 jute 0 kidneybeans 0 lentil ## maize 0 ## mango mothbeans mungbean muskmelon orange ## papaya pigeonpeas ## pomegranate 0 ## 0 4 rice ## watermelon ## predictions kidneybeans lentil maize mango mothbeans mungbean muskmelon ## apple 0 ## banana 0 0 0 0 0 blackgram 0 0 ## chickpea 0 ## 0 0 0 0 coconut coffee 0 ## cotton ## 0 grapes jute ## kidneybeans 0 0 0 ## lentil 30 0 0 0 0 maize 0 30 0 ## mango 30 mothbeans mungbean 0 ## muskmelon 30 orange papaya pigeonpeas pomegranate ## rice watermelon ## predictions orange papaya pigeonpeas pomegranate rice watermelon ## apple ## 0 0 0 0 0 banana 0 ## blackgram 0 0 0 0 0 0 chickpea ## 0 coconut 0 ## coffee 0 0 0 cotton ## grapes ## jute 0 kidneybeans ## lentil 0 0 0 ## maize 0 0 0 0 ## mango mothbeans ## 0 mungbean muskmelon ## orange 30 papaya 30 30 0 pigeonpeas pomegranate 0 30 0 0 rice 0 26 0 watermelon 0 30 model evaluation confusionMatrix(cm) ## Confusion Matrix and Statistics predictions ## apple banana blackgram chickpea coconut coffee cotton grapes jute ## apple 30 0 0 0 0 0 0 banana 30 0 blackgram 0 30 0 0 0 0 ## chickpea 0 0 30 0 0 0 0 coconut ## coffee 0 0 0 30 0 0 ## cotton 0 0 grapes 30 0 ## jute 30 0 ## kidneybeans 0 lentil ## maize 0 ## mango 0 ## mothbeans 0 ## mungbean 0 ## muskmelon ## orange 0 ## papaya 0 pigeonpeas ## pomegranate ## rice ## watermelon ## predictions kidneybeans lentil maize mango mothbeans mungbean muskmelon ## ## apple 0 ## 0 0 0 0 0 0 banana blackgram chickpea 0 0 0 0 ## coconut 0 0 0 0 coffee 0 ## cotton 0 0 ## grapes kidneybeans 0 0 ## lentil 0 0 maize ## 0 0 30 0 mango ## mothbeans 0 0 30 mungbean 0 30 muskmelon 0 30 orange ## 0 papaya ## pigeonpeas 0 pomegranate 0 0 rice watermelon predictions ## ## orange papaya pigeonpeas pomegranate rice watermelon ## apple banana 0 0 blackgram 0 chickpea ## coconut 0 ## 0 0 0 coffee cotton 0 0 grapes jute 0 kidneybeans ## lentil 0 ## maize 0 mango mothbeans 0 0 mungbean 0 muskmelon 0 0 orange 0 30 ## papaya 0 0 pigeonpeas 0 0 0 0 0 0 30 pomegranate 0 0 26 0 30 watermelon Overall Statistics ## ## Accuracy: 0.9924 ## 95% CI: (0.9824, 0.9975) ## No Information Rate : 0.0515

P-Value [Acc > NIR] : < 2.2e-16

Mcnemar's Test P-Value : NA

## Statistics by Class:

## Sensitivity

## Specificity

## Prevalence

Result

## Pos Pred Value

## Neg Pred Value

## Detection Rate

## Detection Prevalence

## Balanced Accuracy

model lies between this range.

## Pos Pred Value

## Neg Pred Value

## Detection Rate

## Detection Prevalence

## Balanced Accuracy

## Pos Pred Value

## Neg Pred Value

## Detection Rate

## Detection Prevalence

## Balanced Accuracy

## Pos Pred Value

## Neg Pred Value

## Detection Rate

## Detection Prevalence

## Balanced Accuracy

## Pos Pred Value

## Neg Pred Value

## Detection Rate

## Detection Prevalence

## Balanced Accuracy

## Pos Pred Value

## Neg Pred Value

## Detection Rate

## Detection Prevalence

## Balanced Accuracy

## Pos Pred Value

## Neg Pred Value

## Detection Rate

## Detection Prevalence

## Balanced Accuracy

Kappa : 0.9921

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

1.00000

0.99842

0.96667

1.00000

0.04394

0.04394

0.04545

0.99921

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

1.00000

1.00000

1.00000

1.00000 0.04545

0.04545

0.04545

1.00000

Overall our model classifies all the crops correctly based on the certain features for the cultivation.

The overall accuracy of the model is 99.24% which means our model correctly classified the observation of dataset.

agreement and and value above 0.8 indicates strong agreement.in our case 0.99 which is suggest the strong agreement.

Class: watermelon

which means no paired sample tested or compared with the analysis.

hypotheis which suggest that our model accuracy is significant.

and detection prevalence (the proportion of predicted positives).

accuracy is 1 which is perfect classification.

Class: apple Class: banana Class: blackgram

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

Class: chickpea Class: coconut Class: coffee Class: cotton

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

0.88235

1.00000

1.00000

0.99365

0.05152

0.04545

0.04545

0.94118

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

Class: muskmelon Class: orange Class: papaya

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545 1.00000

95% confidence interval is a true population parameter. in our case the CI ranges between is 98.24% to 99.75%. which means true accuracy of the

No information rate is a statistical bench mark that is used to evaluate the classification model which means prior probabilities of the most frequent classes in the dataset. The accuracy must be greater than no information which means making meaningful predictions which is useful for practical

Kappa value is the agreement between predicted value and actutal values and it ranges from -1 to 1. in general the value below 0.4 indicates poor

The McNemar's Test is a statistical test used to determine if there is a significant difference between two related proportions. It is often used in cases where the data is paired, such as in a before-and-after study, or when two classifiers are tested on the same dataset in our case its NA

p-value(accuracy > NIR) which is a null hypothesis testing in our case value is(2.2e-16) less than alpha value(0.05) which means reject null

The statistics are also broken down by class. For each class, we have metrics such as sensitivity (the proportion of actual positives that are

positives that are true positives), and negative predictive value (the proportion of predicted negatives that are true negatives). We also have prevalence (the proportion of the data that belongs to each class), detection rate (the proportion of actual positives that are correctly identified),

For example we will take apple the senistivity, specificity, ppv, and npv are all 1 which means our model correctly classifies as a apple. while prevalance and detection rate are equal prevalence proportion is 0.04545 and correctly detected 0.0454 which is a detection rate it means it detected all our proportions correctly detection prevalance is the proportion pf predicted positives which is 0.0454 which is accurate and balanced

correctly identified), specificity (the proportion of actual negatives that are correctly identified), positive predictive value (the proportion of predicted

Class: pigeonpeas Class: pomegranate Class: rice

Class: grapes Class: jute Class: kidneybeans Class: lentil

Class: maize Class: mango Class: mothbeans Class: mungbean

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

1.00000

0.99369 0.86667

1.00000

0.03939

0.03939

0.04545

0.99685

0.96774

1.00000

1.00000

0.99841

0.04697

0.04545

0.04545

0.98387

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

1.00000

1.00000

1.00000

1.00000

0.04545

0.04545

0.04545

1.00000

## ##

## ##

##