BUILDINGASMARTERAI-POWEREDSPAMCLASSIFIER

510521104306: LAVANYA J
PHASE-4:DOCUMENTSUBMISSION

OBJECTIVES:

The problem is to build an AI-powered spam classifier that canaccuratelydistinguishbetweenspamandnon-spammessagesinemailsor text messages. The goal is to reduce the number of false positives(classifyinglegitimatemessagesasspam)andfalsenegatives(missingactualspammessages)whileachievinga highlevelofaccuracy.

PHASE-4:DEVELOPMENTPART-2:

In thispartyouwillcontinuebuildingyourproject.
Inthisphase,we'llcontinuebuildingourspamclassifierby:

- Selectingamachinelearningalgorithm
- Trainingthemodel
- Evaluatingitsperformance.

DATASETLINK:

https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset

ABSTRACT:

Building a smarter AI-powered spam classifier is a compelling solution toaddressthisissue.Thisabstractprovidesanoverviewofthedevelopmentprocessand keyelementsinvolvedincreatinganintelligentandeffectivespamfilter.Thedevelopmentjourneybeginswithdatacollection,encompassingadiversedatasetof spam and non-spam (ham) messages. Model selection is a critical decision,with options ranging from traditional machine learning algorithms to advanceddeep learningarchitectures.

INTRODUCTION:

AsmarterAI-poweredspamclassifierleveragesthecapabilitiesofartificial intelligence, machine learning, and natural language processing to notonly detect and filter spam but also to continually evolve and learn from newthreats. It is an intelligent guardian that ensures that legitimate messages reachtheirintendedrecipientswhilerelegatingunwantedcontenttothedigitalwasteland.Thisdevelopmentjourneyencompassesaseriesofcrucialsteps,eachdesigned to enhance the classifier's efficacy. It begins with the collection of adiversedatasetcontainingexamplesofspamandlegitimatemessages.

Data preprocessing tasks prepare this data for model training, includingtext normalization and feature extraction. Model selection is a pivotal decision,where various machine learning algorithms or deep learning architectures areconsidered.

SELECTINGANMACHINELEARNINGALGORITHM

ListofPopularMachineLearningAlgorithm

1. **LinearRegressionAlgorithm**
2. **LogisticRegressionAlgorithm**
3. **DecisionTree**
4. **SVM**
5. **NaïveBayes**
6. **KNN**
7. **K-MeansClustering**

1. LinearRegression

Linear regression is one of the most popular and simple machine learningalgorithms that is used for predictive analysis. Here, **predictive analysis** definesprediction of something, and linear regression makes predictions for *continuousnumbers*suchas**salary,age,etc.**

It shows the linear relationship between the dependent and independentvariables, and shows how the dependent variable(y) changes according to theindependentvariable(x).

It tries to best fit a line between the dependent and independent variables,andthisbestfitlineisknownsastheregressionline.

Theequationfortheregressionlineis:

$y=a_0+a*x+b$

Here,y=dependentvariable

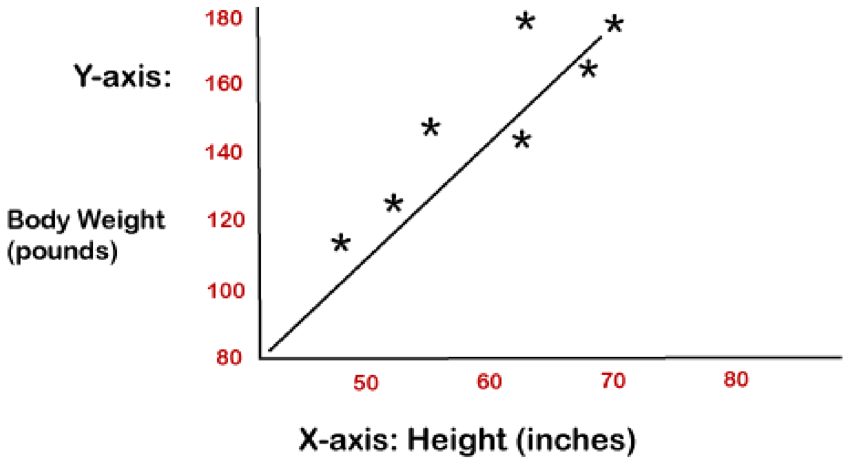
x=independentvariablea

$_0$=Interceptofline.

Linearregressionisfurtherdivided intotwotypes:

- **SimpleLinearRegression:** Insimplelinearregression,asingleindependentvariableisusedtopredictthevalueofthedependentvariable.
- **MultipleLinearRegression:**Inmultiplelinearregression,morethanoneindependentvariablesareusedtopredictthevalueof thedependentvariable.

Thebelowdiagramshowsthelinearregressionforpredictionofweightaccordingtoheight:



2. LogisticRegression

Logisticregressionisthesupervisedlearningalgorithm,whichisusedto**predictthecategoricalvariablesordiscretevalues**.Itcanbeusedforthe *classification problems in machine learning,* and the output of the logisticregressionalgorithmcanbe eitherYesorNO,0 or1,Red orBlue,etc.

Logistic regression is similar to the linear regression except how they areused, such as Linear regression is used to solve the regression problem and predictcontinuousvalues,whereasLogisticregressionisusedtosolvetheClassificationproblemandused to predictthediscretevalues.

Instead of fitting the best fit line, it forms an S-shaped curve that liesbetween 0 and 1. The S-shaped curve is also known as a logistic function thatuses the concept of the threshold. Any value above the threshold will tend to 1,and belowthethresholdwilltend to0.

3. DecisionTreeAlgorithm

A decision tree is a supervised learning algorithm that is mainly used tosolve the classification problems but can also be used for solving the regressionproblems. It can work with both categorical variables and continuous variables.Itshowsatree-likestructurethatincludesnodesandbranches,andstartswiththerootnodethatexpandonfurtherbranchestilltheleafnode.The**internalnode**isused to represent the **features of the dataset, branches show the decisionrules,**and **leafnodesrepresenttheoutcome ofthe problem.**

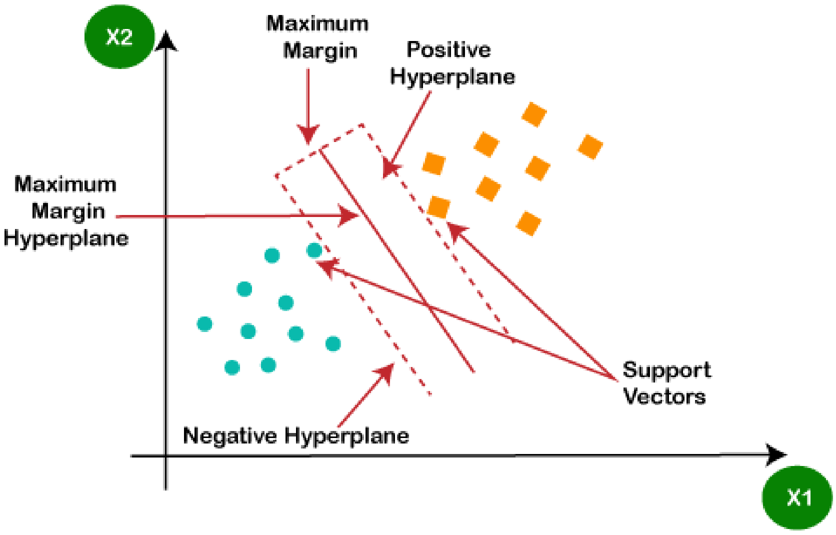
Somereal-worldapplicationsofdecisiontreealgorithmsareidentificationbetween cancerous and non-cancerous cells, suggestions to customers to buy acar,etc.

4. SupportVectorMachineAlgorithm

A support vector machine or SVM is a supervised learning algorithm thatcan also be usedfor classification and regression problems. However,it isprimarily used for classification problems. The goal of SVM is to create ahyperplaneordecisionboundarythatcansegregatedatasetsintodifferentclasses.

The data points that help to define the hyperplane are known as **supportvectors**,andhenceitisnamedassupportvectormachine algorithm.

Somereal-lifeapplicationsofSVMare **facedetection,imageclassification,Drugdiscovery**,etc.Considerthe below diagram:



Aswecanseeintheabovediagram,thehyperplanehasclassifieddatasetsinto twodifferentclasses.

5. NaïveBayesAlgorithm:

NaïveBayesclassifierisasupervisedlearningalgorithm,whichisusedtomake predictions based on the probability of the object. The algorithm named asNaïve Bayes as it is based on **Bayes theorem**, and follows the *naïve* assumptionthat says'variablesareindependent ofeach other.

The Bayes theorem is based on the conditional probability; it means thelikelihood that event(A) will happen, when it is given that event(B) has alreadyhappened.Theequation forBayestheorem isgivenas:

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

Naïve Bayes classifier is one of the best classifiers that provide a goodresult for a given problem. It is easy to build a naïve bayesian model, and wellsuitedforthehugeamountofdataset.Itismostlyusedfor**textclassification**.

6. K-NearestNeighbour(KNN)

K-Nearest Neighbour is a supervised learning algorithm that can be usedforbothclassificationandregressionproblems.Thisalgorithmworksbyassuming the similarities between the new data point and available data points.Based on these similarities, the new data points are put in the most similarcategories. It is also known as the lazy learner algorithm as it stores all theavailable datasets and classifies each new case with the help of K-neighbours.The new case is assigned to the nearest class with most similarities, and anydistance function measures the distance between the data points. The distancefunction can be **Euclidean, Minkowski, Manhattan, or Hamming distance**,based ontherequirement.

7. K-MeansClustering

K-meansclusteringisoneofthesimplestunsupervisedlearningalgorithms, which is used to solve the clustering problems. The datasets aregrouped into K different clusters based on similarities and dissimilarities, itmeans, datasets with most of the commonalties remain in one cluster which hasverylessornocommonalitiesbetweenotherclusters.InK-means,K-referstothenumber of clusters, and **means** refer to the averaging the dataset in order to findthecentroid.

It is a centroid-based algorithm, and each cluster is associated with a centroid.This algorithm aims to reduce the distance between the data points and theircentroidswithinacluster.

This algorithm starts with a group of randomly selected centroids that form theclusters at starting and then perform theiterative process to optimize thesecentroids'positions.

Itcanbeusedforspamdetectionandfiltering,identificationoffakenews,etc.

TRAININGMODELS:

- DataCollection:Gatherasubstantialanddiversedatasetofbothspamandnon-spam messages. This dataset shouldencompassvariousforms ofcommunicationlikeemails, textmessages,andsocialmediacontent.

- Data Preprocessing: Clean the data by removing irrelevant information(e.g.,emailheaders,formatting)andstandardizingthetext.Textpreprocessing tasks may include tokenization, stemming, and removingspecialcharactersor stop words.

- Feature Extraction: Convert the text data into numerical features that themachine learning model can understand. Common techniques include TF-IDF(TermFrequency-InverseDocumentFrequency)andwordembeddings(e.g.,Word2VecorGloVe).

- Labeling: Annotate the dataset, marking each message as either spam ornon-spam.Ensurethedatasetiswell-balancedtoavoidbias.ModelSelection: Choose an appropriatemachine learning algorithm or deeplearning architecture for your spam classifier. Common choices includedecisiontrees,randomforests,supportvectormachines,orneuralnetworks.

- SplitData:Dividethedatasetintotraining,validation,andtestingsets.Thetraining set is used to teach the model, the validation set to fine-tunehyperparameters,andthetestingsettoevaluatethemodel'sperformance.

- Model Training: Train the selected model using the training data. Themodellearnstorecognizepatternsandfeaturesthatdistinguishspamfromnon-spam messages.

- HyperparameterTuning:Experimentwithdifferenthyperparameters(e.g.,learning rates, batch sizes, number of layers) to optimize the model'sperformance. Use the validation set to assess the model's performanceduringthisprocess.

- Evaluation: Evaluate the model's performance on the test dataset usingvariousmetricslikeprecision,recall,F1-score,andaccuracy.Thesemetrics helpmeasurethemodel'sabilitytocorrectlyclassifyspam

EVALUATING THE PERFORMANCE OF A SMARTER AI-POWEREDSPAMCLASSIFIER

- Confusion Matrix: Create a confusion matrix to visualize the classifier'sperformance. It categorizes results into four groups: true positives, truenegatives,falsepositives,andfalsenegatives.

- Accuracy: Calculate accuracy by dividing the sum of true positives andtrue negatives by the total number of examples. However, accuracy alonecanbemisleading,especially in imbalanced datasets.

- Precision: Precision measures the proportion of true positive predictionsamong all positive predictions. It helps determine how often the classifiercorrectlyidentifiesspamwithoutfalselylabelingnon-spamas spam.

- Recall(Sensitivity):Recallcalculatestheproportionoftruepositivepredictions among all actual positive cases. It shows how well the classifiercapturesallspam messageswithoutmissingtoomany.

- F1-Score:TheF1-scorecombinesprecisionandrecallintoasinglemetric,which is useful when you want to balance the trade-off between falsepositivesandfalsenegatives.

- Specificity:Specificitymeasurestheproportionoftruenegativepredictions among all actual negative cases. It's crucial to assess how welltheclassifier avoidsfalsely labeling non-spamasspam.

- ROC Curve and AUC: Plot the Receiver Operating Characteristic (ROC)curvetovisualizetheclassifier'sperformanceatdifferentthresholdlevels.The Area Under the Curve (AUC) quantifies the overall performance. AhigherAUCindicatesbetterperformance.

- Cross-Validation:Usek-foldcross-validationtoassessthemodel'srobustness and generalization across different subsets of the data. Thishelpsavoidoverfittingandprovidesamoreaccurateestimateofperformance.

- Precision-Recall Curve: Plot a precision-recall curve to understand howprecisionandrecallchangeatvariousthresholdlevels.Thisisparticularlyusefulwhendealingwithimbalanced datasets.

- FalsePositiveRate(FPR):CalculateFPRastheproportionoffalsepositivestoactualnegatives.It'sessentialtoensurethatnon-spammessagesarenotfrequently misclassified asspam.

- False Negative Rate (FNR): Measure FNR asthe proportion of falsenegativestoactualpositives.It'simportanttominimizetheriskofmissinggenuinespammessages.

- User Feedback: Gather feedback from users to identify false positives(legitimatemessagesclassifiedasspam)andfalsenegatives(spammessagesthatbypassthefilter).Thisfeedbackcanguidemodelimprovements.

- Bias and Fairness Analysis: Evaluate the model for bias and fairness toensure that it doesn't disproportionately impact specific user groups. Usefairnessmetricslikedisparateimpactandequalopportunitytoassessthis.

- A/B Testing: Conduct A/B testing by deploying the model in a controlledmannerandcomparingitsperformancetothepreviousssystemorversions.