**Name**: Rudra Swant

**Roll no**: 282074

**Batch**: B3

**Assignment 7**

# ProblemStatement:

EveryyearmanystudentstaketheGREexamtogetadmissioninforeignUniversities.The data set contains GRE Scores (out of340),TOEFLScores(outof120),UniversityRating (out of 5), Statement ofPurposestrength(outof5),LetterofRecommendationstrength (outof5),UndergraduateGPA(outof10),ResearchExperience(0=no,1=yes),Admitted (0=no,1=yes).Admittedisthetargetvariable.

DataSet:<https://www.kaggle.com/mohansacharya/graduate-admissions>

Thecounselorofthefirmissupposedtocheckwhetherthestudentwillgetanadmission ornotbasedonhis/herGREscoreandAcademicScore.Sotohelpthecounselortomake appropriate decisions build a machine learning model classifier using a Decision tree to predictwhetherastudentwillgetadmissionornot.

1. ApplyDatapre-processing(LabelEncoding,DataTransformation….)techniques if necessary.
2. Performdatapreparation(Train-TestSplit)
3. ApplyMachineLearningAlgorithm
4. EvaluateModel.

# Objectives:

* 1. Topreprocessandcleanthedatasetforimprovedmodelperformance.
  2. Tosplitthedatasetforeffectivemodeltrainingandevaluation.
  3. TobuildandapplyaDecisionTreeclassifierforadmissionprediction.
  4. Toassessthemodel’saccuracyandeffectiveness.

# Resourcesused:

1. Softwareused:VisualStudioCode
2. Librariesused:Pandas,Matplotlib,sklearn,Seaborn

# Theory:

1. **DecisionTreeClassifier**

ADecisionTreeisasupervisedlearningalgorithmusedforclassificationandregression tasks.Itsplitsthedatasetintosmallersubsetsusingconditionsbasedonfeaturevalues, formingatree-likestructure.ThemaincomponentsofaDecisionTreeare:

* + **RootNode:**Representstheentiredatasetandselectsthebestfeatureforsplitting.
  + **InternalNodes:**Representdecisionpointsbasedonattributevalues.
  + **LeafNodes:**Representfinalclassificationlabels(Admission=Yes/No).

Themodelworksbyrecursivelysplittingthedatabasedonfeaturesthatresultinthe highestinformationgainorlowestGiniimpurity.

1. **SummaryStatistics**

Summarystatisticsprovidekeyinsightsintothedataset,including:

* + **Minimum&MaximumValues:**Identifydatarange.
  + **Mean&Median:**Representcentraltendency.
  + **StandardDeviation&Variance:**Measuredatadispersion.
  + **Percentiles:**Showdatadistribution.

1. **DataPreprocessing**

Datapreprocessingensuresthedatasetiscleanandreadyforanalysis.Stepsinclude:

* + Handlingmissingvalues(e.g.,imputation,removal).
  + Encodingcategoricalvariables(e.g.,LabelEncoding,One-HotEncoding).
  + Normalizationandscalingfornumericalstability.

1. **ModelEvaluationMetrics**

Toassessmodelperformance,weuse:

* + **Accuracy:**Measurescorrectpredictionsovertotalinstances.
  + **Precision&Recall:**Evaluatepositiveclassperformance.
  + **F1-score:**Balancesprecisionandrecall.
  + **ConfusionMatrix:**Displaystrue/falsepositivesandnegatives.

# Methodology:

1. **DataPreprocessing**
   * Handlemissingvalues,ifany,usingimputationtechniques.
   * Performfeaturescalingornormalization,ifrequired.
   * Encodecategoricalfeaturesifnecessary.
2. **DataPreparation**
   * Splitthedatasetintotraining(80%)andtesting(20%)subsetsusingScikit-Learn’s train\_test\_split()function.
3. **ModelImplementation**
   * ImplementaDecisionTreeClassifierusingScikit-Learn’sDecisionTreeClassifier.
   * Trainthemodelusingthetrainingdataset.
   * Predictadmissionoutcomesusingthetestdataset.
4. **ModelEvaluation**
   * EvaluatemodelperformanceusingmetricssuchasAccuracy,Precision,Recall,and F1-score.
   * UseaConfusionMatrixtoanalyzemodelpredictions.

# Conclusion:

* + Preprocessingensuresthatthedatasetiscleanandsuitableformodeling.
  + ADecisionTreeclassifiercaneffectivelyclassifystudentsbasedontheiracademic scores.
  + Evaluatingperformancemetricshelpsinunderstandingmodelaccuracyand effectiveness.