Etep ! Proport paudos as pol Proposit matiplotlyb, pyphot as plt. frain-data: pd. road-csv(1 train-loan.csv) train-data. head() Avair data. Shape train-data columns from - data, dtypes train-data. Pofo() Frain -data value \_eounts train-data topendants. Value - Counts () #Replace numbers os - string by . Pritagex . in Dependants . Columns def shing (x): if X = = 101: return 'bad' elf. x== 111: return 'average' elif x== 121: return igood! else. setus rescellant frain-dola ['Despardants'] = train-data ['Depardants'] apply (Anny) train\_data. Pena(), Sum()

Scanned with CamScanner

## Lab7. Loan Approval Classification using SVM

#### Objectives

in this lab, you will build a classification model to classify the loan applicants into eligible applicants or not eligible applicants using Support Vector Machine.

### Learning Outcomes

After completing this lab, you will be able to

- · Apply data cleaning methods
- Perform EDA and understand who got their loans approved
- Do feature engineering with One Hot Encoding
- Create LinearSVC model, train and predict on the data
- Print accuracy, confusion matrix and classification report
- Compare LinearSVC model with SVC and SGDClassifier models

#### Business Use Case

Heber Housing Finance deals in all home loans. They have presence across all urban, semi urban and rural areas. Customer first apply for home loan after that company validates the customer eligibility for loan. However doing this manually takes a lot of time. Hence, it wants you to automate the loan approval process (real time) based on customer information. So you should identify all features and build a model that enable the company to approve the load application or not.

The dataset contains the details of 614 loan applicants, where each applicant is described with 12 features. Loan\_Status is the target feature (ie., dependent variable) and all others are independent variables.

Step1. [Understand Data]. Using Pandas, import "train\_loan.csv" file and print properties such as head, shape, columns, dtype, info and value\_counts.

### Step 2. [Data Cleaning]

- Replace numbers as string by integer in "Dependents" column
- Fill missing data in categorical columns (Gender, Married, Dependents, Education, Self Employed, Credit\_History) by its mode value
- Handle missing values in numberical columns
- Drop Loan ID column

### Step3. [OPTIONAL: Exploratory Data Analysis - Who got their loan approved]

Draw count plot for

- Married?
- Dependents?
- Graduates?
- Self-employed?

### Step4. [Extract X and y] from the dataframe

### Step5. [One Hot Encoding]

Perform OHE or categorical columns, use this method: X = pd.get\_dummies(X)

### Step6. [Model Building]

- 3

•

- Split X and y for training and testing
- Using StandardScaler, fit\_transform on X\_train and transform on X\_test values
- create LinearSVC model, train and test
- print accuracy value
- Print confusion matrix between y\_test and y\_pred
- Print dassification\_report

DESCRIPTION OF THE PROPERTY OF

```
# Cortegorical
  town-data ['Grender] of All na (tain_data [Grenders]). mode() (o),
                                    Proplace = True)
  tourn-data [Harried]. fill ra (tourn-data ('Harried)]. model) []
                                 Proplace strue)
  train-data ('Depardents!) till na (train-data ('Depardents).
                                              mode () [o], implace Tow)
   food - data [ Lour-Amount-Tam] fill not both data [ Lour-Amount-leads
                                       ·mode()[0], Implace=True
    fram.data [ 'Credit - History Tofil no (frain-data [ credit History ].
                                    model) (0), Inplace = True)
  train-data ['Self-Employed'] tillna (train-data['Self-Employed,
                                  imode()[o], implace=True)
   numerical
    train data [ Lean Amount ] of the rate carn data [ toan Amount ].
train data = train - data drop [ [ fran D], axis=) mean (). Inplace=True)
 pit-subplof231)
 train-data [ Mounted]. Value-Counts (). Plot ( Lind = 1 box),
                                                  fitte = Gorder,
PH-Subplot (232)
                                                  figsize = (2010)
   Frain-dota [ Dependents] - value - counts (). plot (49 nd = 1 boar)
                                                title = Dependents!)
 pH. Subplot (233)
      from -data [ Education] . value counts (). plot (kind = bar)
                                               title = 'Education')
  Plt. Subplot (234)
      train = data [ Self-Employed ]. Value - Counts ().
        Af-show()
```

# Step7. [Performance Comparisons]

- Compare the performance of LinearSVC against LogisticRegression
- Compare LinearSVC against SVC with various kernels such as 'linear', 'poly', 'rbf' and 'sigmoid'

X = Loan\_data.doop(['Loan\_status])]/axis=]

Y = train\_data. Loan\_status

Propost warrings

Mary last filtermanings (1 Egrore)

X= & pd. get\_dummes(x)

drom stream. model - selection Import train - text-split x\_town, x\_test, y-town, y\_test = train\_test\_split(x,y)

test\_99ze=0.3, random\_state=0)

from skleam. preprocessing Pomport Standard Scalar

Scale = standard Scalar()

SS = Scale. Let. Assansform (x-frain)

SSI = Scale. transform (x-test)

drom skilearn. Stm import finear Stc.

model = Linear &vc()

mode). fit (ss, y-tookn)

LSVC-y-pred = model. predict(ses)

Lave-u- pred

from Iteam, meters Proport accusacy\_score, befulon-nation, Classification\_report accuracy-score(y-test, Lsvc-y-pred)

DEPT OF DATA SCIENCE I BISHOP HEBER COLLEGE I TRICHY

DR.K.RAJKUMAR

```
fred book
        8 fb = 9ve = Sum - Svo (Excesse) 51 xbb', C=10)
         879-265-724 (CCV, FRUIN)
          27 feach and - dusq = 478- 216- boospert (33)
        ergnord. Sic = Erm. Cre ( Ferre! = 1 stopmord', C=1.0)
          ergmord se by (a, y train)
          64 Sigse-9-poed = signord-sic. profict (se 1)
         PSPA("Linear . Svc: ", accuracy score (g. feet, Inc. yaxed))
          ARA (Alfoly-Sic: "occuracy-store (y-test, psic-y-pred))
           print ("alt suc", accuracy-score (y-toot, ybbeve-y-pred)
            paint "Signord suc:", accuracy-score (y-test, signo-y-pred) [
      Modelses = 13
        modalses. append (('sve', Linearsve (1))
        modelses. append ( ( 'sve . boly', sum. Sve (kexnel= poly,
                                                                       6
                                                  C=1.0)))
          moleless. append(('suc'rbb', sums vc(korne) = 1.40b',
          modelsss-append(('se poly', svm. swe/texnel1 = 1 sigmoid',
                                                   c = (.0)))
          results = [ ]
          names = []
          scoring = laccuracy
           for mane, madel in model 35%!
         kfuld = mother-selection. Good (n_sple) =10)
          Ctresuffe = model - selection. Cross val - eare (mode),
            results. append (cv-rosulte)
               names. append (cv-routh)
               rancs, append (rane)
                prof [mag)
```

NOTES

Conforsion\_matrix (y-test, Levery-pred)

Point (classification-report (y-test, Love y-poed))

doom Sictean. (mear-model Propost - Logistic Ragnession

1gr = Kogestic Regession ()

1gr. fit (ss, y-train)

lr. y pred = 1gs. predict(ss 1)

from . Sklearn. Stm Impost Fineau EVC.

1\_Ste = Linear Ste()

1-svc. fit (:cs, /-train)

Isve-y-pred =1-sve-predet (1811)

Porting ("Loggistic Regression:", accuracy-score(y-test, by-y-pood)

Point (" Linear SVC: ", accuracy - Store (x-test, 18/c-y-pred)

from siteasn import sum, model\_selection

model=[]

model append f('Sve'; Lineau Sve()))

model. appead (('LR', LogPatha Regression ()))

vesults = []

names = []

Storing = 1 accuracy

for poure, model in model

Kford = madel = Selection, kgold (n-9phits = 10)

EPT OF DATA SCIENCE | BISHOP NEBER COLLEGE | TRICHY

CV-VOSUITO = modelo - selection. Cross-Val \_ score (model, x train, V train, Scoring) result append (cv-result) hames append (name) mag = 110/05: 0/0/6. (0/06)".0/0 (name, ev-vesul &-mean (), CV xesults sld() point (mag) fly = blt. Byuse () fig. Suplite (Compartson between different MLASI) .ax = flg. add\_subplot(11) px. boxplot (results) plt-show() from Schear. Chear-model Proport SBDC/ essifier modelSS=[] mode is appeard (['SVC', HREAY SUCCO)) models append (('sout, scide lassifical")) results=[] rames [7 Gorfra = 'accuracy' for name, model in models! Kfold = model- selection. brold (n\_splits=10) CV\_results=mode)\_selection. crosseval\_score (model, X-train, X-touth, CV=26/pbd, Scoring. results append (cv-results) hames append (name) mag= " 908: "90", (9/0)" offrame, ex- results. mean (), ev-results, stall) point (meg)

Scanned with CamScanner

**NOTES** ·fig=plt.figure() deg. Suptitle (1 Compaisison between different. MLAS) ax = fig. add-subplot(111) Plf-boxplol (results) Plt-show () from skleam. linear\_model. import-scitclassifiex egd = 16Ddassifica() sqd. Let (ss, y-train) s gdc-y-pood = sqd. poodict (sed) from Skleasn, sum. Propost linear suc 1-svc= Linear Svc () 1 : svc- LP+ (ss, y-toon) 1 src-4-pred = 1-svc. predict(ss1) pant ("scroelassaffer;", accuracy-score (y-test, sack-ytsel)) print ("Lineas eve: ", accuracy score (y-test, Isve-y-pred))

3). Doom Sklearn. SVM. Pompord SVC

1. Src = [Preassic()

1 -src. for (ss, y -town)

1 svc\_4-pred = 1 -svc.predict(ss1)

poly-suc-sum-suc (ternel = 1 poly), (=1.0)

price y-pred = poly ovc. predict (SS1)

DEPT OF DATA SCIENCE | BISHOP HEBER COLLEGE | TRICHT

DR.K.RAJKUMAR