IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

Travel Insurance Prediction

Vanam Sanjana[1], Shaik Noorul Huda[2], Dr. Ch Surya Kiran[3]
IVth Year Students Computer Science and Engineering [1,2]
Associate Professor, Department of Computer Science and Engineering[3]
NRI Institute of Technology ,Vijayawada, AP, India

Abstract: *Insurance is a contract whereby an individual obtains financial protection against losses from an insurance company against the risks of financial losses as mentioned in the insurance. Many types of insurance exist today and there are so many companies that offer insurance services. These companies always need to predict whether or not a person will buy insurance so that they can save time and money for the most profitable customers. The task of insurance prediction in which the travel insurance prediction can be taken into progress using the machine learning using python. We here get the data from the databases from the companies. Here our task is to train a machine learning model to predict whether an individual can purchase the insurance policy from the company or not. We also use CSV (comma separated value) files in the process.*

Keywords: python, visualization, ensembling, regression, warnings module

I. INTRODUCTION

In case you are travelling abroad, a travel insurance policy protects you against losses suffered due to loss of baggage, delays in flight and trip cancellation. Travelers need to read their insurance policies cautiously to perceive the covered aspects and to detect prohibitions, if any. It is also integral for those working abroad to purchase a travel insurance policy. Travel Insurance is an important purchase for consumers when it comes to travel planning. Travel Insurance expenditure is heavily linked to consumer requirements. Based on personal requirements of their age, health status, family demographics, travel frequencies, consumers select the options for Travel Insurance products available on the market to suit their needs. Analyzing current datasets based on factors such as age groups, employment status, Annual Income amounts, education levels, health statuses, family backgrounds helps to provide information on their spending patterns. Models can be generated to predict consumer spending trends. Such trends help provide Travel Insurance providers design better products with suitable customizable options. This can better cater to consumer needs thereby improving service efficiency and quality levels

II. PROPOSED SYSTEM

For the further development of travel insurance prediction, we can join this data frame to other data, to create complete data set. More modelling can be performed with improved accuracy obtained. Forwarded by applying the machine learning tasks like data exploration, data cleaning, feature engineering and model building. It is more accurate using the other data in the process. For example if we are calculating insurance based on a person age, health and salary them we can get the accuracy by decision tree method but it is not up to the mark. But if we add another data like frequent flyers then we can get more accurate results that is only the frequent flyers would be provided insurance and others would not get it. So the company also would not loose money and they can have good prediction regarding the insurance buyers and non-buyers

III. TECHNOLOGIES USED

- **Python:** Python is commonly used for developing websites and software, task automation, data analysis, and data visualization. Since it's relatively easy to learn, Python has been adopted by many non-programmers such as accountants and scientists, for a variety of everyday tasks, like organizing finances.
- **Visualization:** The purpose of data visualization is pretty clear. It is to make sense of the data and use the information for the organization's benefits. That said, data is complicated, and it gains more value as and when

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

it gets visualized. Without visualization, it is challenging to quickly communicate the data findings and identify patterns to pull insights and interact with the data seamlessly.

- **Matlab:** MATLAB has gained wide recognition because of its importance in floatingpoint linear algebra. As far as trading is concerned, this programming language can generate plots and other such interactive tasks which is why it is favored by a majority of the traders out there.
- **Numpy:** Numpy or Numerical Python, provides powerful implementations of large multi-dimensional arrays and matrices. The library consists of functions for complex array processing and high-level computations on these arrays

Software Requirements:

- Browser like Google Chrome
- Visual Studio Code Editor
- PgAdmin4 for PostgreSQL database
- Python 3.7 Django Framework
- Alpaca Trading and Paper Trading Account

Hardware Requirements:

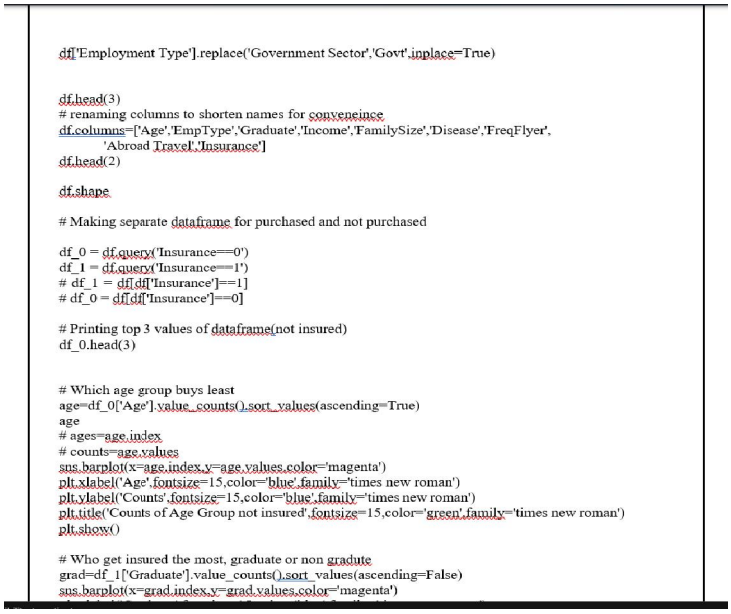
- Inter Core i7-9750H
- CPU 2.60GHz
- 8GB RAM
- 1TB HDD

IV. CODING

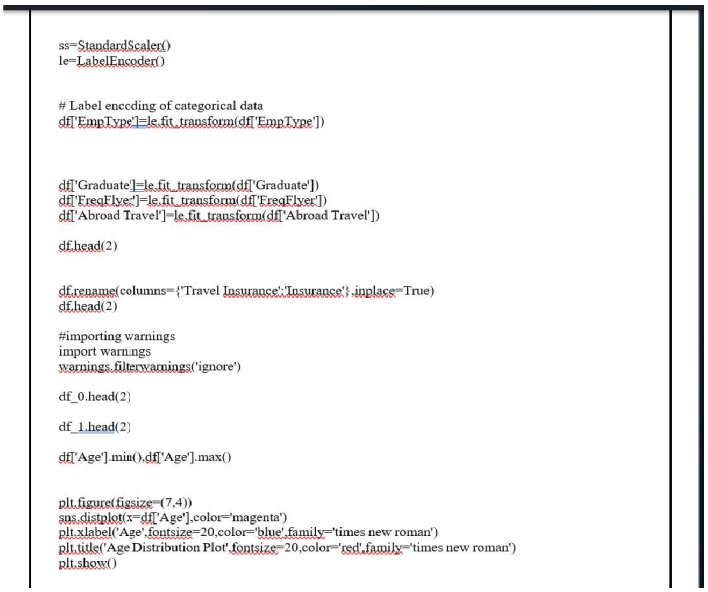
5. CODING

```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
%matplotlib inline


import pandas as pd
data = pd.read_csv('TravelInsurancePrediction.csv')
data.head()

df=data.copy()

df.shape

del df['Unnamed: 0']

df.shape

textcol=df.select_dtypes(include='object').columns
floatcol=df.select_dtypes(exclude='object').columns

for i in textcol:
    print(i,'\n',df[i].unique())
for i in floatcol:
    print(i,'\n',df[i].unique())


# checking for null values
df.isnull().sum()
```

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023



```
df['Employment Type'].replace('Government Sector','Govt',inplace=True)


df.head(3)
# renaming columns to shorten names for convenience
df.columns=['Age','EmpType','Graduate','Income','FamilySize','Disease','FreqFlyer',
            'Abroad Travel','Insurance']
df.head(2)

df.shape

# Making separate dataframe for purchased and not purchased

df_0 = df.query('Insurance==0')
df_1 = df.query('Insurance==1')
# df_1 = df[df['Insurance']==1]
# df_0 = df[df['Insurance']==0]

# Printing top 3 values of dataframe(not insured)
df_0.head(3)


# Which age group buys least
age=df_0['Age'].value_counts().sort_values(ascending=True)
age
# ages=age.index
# counts=age.values
sns.barplot(x=age.index,y=age.values,color='magenta')
plt.xlabel('Age',fontsize=15,color='blue',family='times new roman')
plt.ylabel('Counts',fontsize=15,color='blue',family='times new roman')
plt.title('Counts of Age Group not insured',fontsize=15,color='green',family='times new roman')
plt.show()

# Who get insured the most, graduate or non gradute
grad=df_1['Graduate'].value_counts().sort_values(ascending=False)
sns.barplot(x=grad.index,y=grad.values,color='magenta')
```



```
ss=StandardScaler()
le=LabelEncoder()


# Label encoding of categorical data
df['EmpType']=le.fit_transform(df['EmpType'])



df['Graduate']=le.fit_transform(df['Graduate'])
df['FreqFlyer']=le.fit_transform(df['FreqFlyer'])
df['Abroad Travel']=le.fit_transform(df['Abroad Travel'])

df.head(2)


df.rename(columns={'Travel Insurance':'Insurance'},inplace=True)
df.head(2)

#importing warnings
import warnings
warnings.filterwarnings('ignore')

df_0.head(2)

df_1.head(2)

df['Age'].min(),df['Age'].max()


plt.figure(figsize=(7,4))
sns.distplot(x=df['Age'],color='magenta')
plt.xlabel('Age',fontsize=20,color='blue',family='times new roman')
plt.title('Age Distribution Plot',fontsize=20,color='red',family='times new roman')
plt.show()
```

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

```
plt.xlabel('Age',fontsize=20,color='blue',family='times new roman')
plt.title('Age Distribution Plot',fontsize=20,color='red',family='times new roman')
plt.show()


age=df.Age.value_counts()
a=age.index
b=age.values


plt.figure(figsize=(10,4))
sns.barplot(x=a,y=b,color='green')
plt.xlabel('Age',fontsize=20,color='blue',family='times new roman')
# plt.ylabel('Insurance',fontsize=20,color='blue',family='times new roman')
plt.title('Age Distribution',fontsize=20,color='blue',family='times new roman')
plt.show()


emp=df.EmpType.value_counts()
```

12

```
a=emp.index
b=emp.values


plt.figure(figsize=(2,4))
sns.barplot(x=a,y=b,color='red')
plt.xlabel('Employment type',fontsize=20,color='blue',family='times new roman')
# plt.ylabel('Insurance',fontsize=20,color='blue',family='times new roman')
plt.title('Employement Count',fontsize=20,color='blue',family='times new roman')
```

```
x=df.drop('Insurance',axis=1)
y=df.Insurance


x=ss.fit_transform(x)

#split data into training and testing data set
#train model on training data set
#test model on testing data set
from sklearn.model_selection import train_test_split


xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.25,random_state=1)


# Logistic Regression Model
from sklearn.linear_model import LogisticRegression
model1=LogisticRegression(fit_intercept=True,C=1)


model1.fit(xtrain,ytrain)
model1.score(xtrain,ytrain),model1.score(xtest,ytest)


# Adaboost Classifier
from sklearn.ensemble import AdaBoostClassifier,RandomForestClassifier
abc=AdaBoostClassifier(n_estimators=50)
rfc=RandomForestClassifier(n_estimators=50)


abc.fit(xtrain,ytrain)
abc.score(xtrain,ytrain),abc.score(xtest,ytest)


rfc.fit(xtrain,ytrain)
rfc.score(xtrain,ytrain),rfc.score(xtest,ytest)
```


DOI: 10.48175/IJARSCT-11260

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

V. DESCRIPTION

```
import pandas as pd
data = pd.read_csv('TravelInsurancePrediction.csv')
data.head()
```

	Unnamed: 0	Age	Employment Type	GraduateOrNot	AnnualIncome	FamilyMembers	ChronicDiseases	FrequentFlyer	EverTravelledAbroad	TravelInsurance
0	0	31	Government Sector	Yes	400000	6	1	No	No	0
1	1	31	Private Sector/Self Employed	Yes	1250000	7	0	No	No	0
2	2	34	Private Sector/Self Employed	Yes	500000	4	1	No	No	1
3	3	28	Private Sector/Self Employed	Yes	700000	3	1	No	No	0
4	4	28	Private Sector/Self Employed	Yes	700000	8	1	Yes	No	0

Figure 1: It shows importing the required libraries and uploading the csv file

```
df.head(3)
```

	Age	Employment Type	GraduateOrNot	AnnualIncome	FamilyMembers	ChronicDiseases	FrequentFlyer	EverTravelledAbroad	TravelInsurance
0	31	Govt	Yes	400000	6	1	No	No	0
1	31	PrivateS ctr	Yes	1250000	7	0	No	No	0
2	34	PrivateS elf	Yes	500000	4	1	No	No	1

```
# renaming columns to shorten names for conveneince
df.columns=['Age','EmpType','Graduate','Income','FamilySize','Disease','FreqFlyer',
            'Abroad Travel','Insurance']
df.head(2)
```

	Age	EmpType	Graduate	Income	FamilySize	Disease	FreqFlyer	Abroad Travel	Insurance
0	31	Govt	Yes	400000	6	1	No	No	0
1	31	PrivateS elf	Yes	1250000	7	0	No	No	0

Figure 2:it shows the annual income of govt employees

```
textcol=df.select_dtypes(include='object').columns
floatcol=df.select_dtypes(exclude='object').columns

for i in textcol:
    print(i,'\n',df[i].unique())
for i in floatcol:
    print(i,'\n',df[i].unique())
```

```
Employment Type
 ['Government Sector' 'Private Sector/Self Employed']
GraduateOrNot
 ['Yes' 'No']
FrequentFlyer
 ['No' 'Yes']
EverTravelledAbroad
 ['No' 'Yes']
Age
 [31 34 28 25 33 26 32 29 35 30 27]
AnnualIncome
 [ 400000 1250000  500000  700000 1150000 1300000 1350000 1450000  800000
 1400000  850000 1500000 1050000  350000 1100000  600000  900000  550000
  300000  750000 1200000 1000000  950000 1700000 1750000  650000  450000
 1650000 1800000 1550000]
FamilyMembers
 [6 7 4 3 8 9 5 2]
ChronicDiseases
 [1 0]
TravelInsurance
 [0 1]
```

Figure 3:It shows the complete details of the family

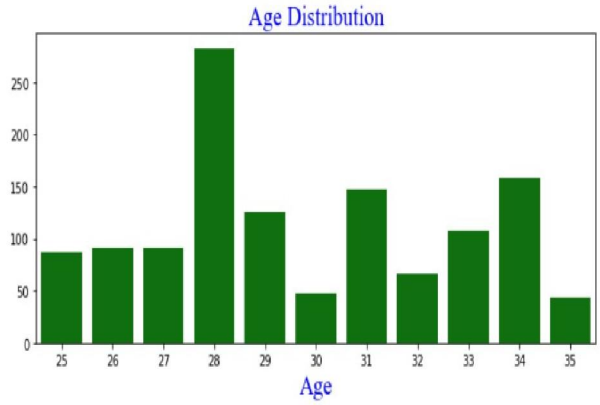
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

```
plt.figure(figsize=(7,4))
sns.distplot(x=df['Age'],color='magenta')
plt.xlabel('Age',fontsize=20,color='blue',family='times new roman')
plt.title('Age Distribution Plot',fontsize=20,color='red',family='times new roman')
plt.show()
```


```
age=df.Age.value_counts()
a=age.index
b=age.values
```

Figure 4: It plots the data we collected

```
plt.figure(figsize=(10,4))
sns.barplot(x=a,y=b,color='green')
plt.xlabel('Age',fontsize=20,color='blue',family='times new roman')
# plt.ylabel('Insurance',fontsize=20,color='blue',family='times new roman')
plt.title('Age Distribution',fontsize=20,color='blue',family='times new roman')
plt.show()
```




```
emp=df.EmpType.value_counts()
a=emp.index
b=emp.values
```

Figure 5: Final distribution of data based on ages

VI. CONCLUSION

The task of insurance prediction is something that adds value to every insurance company. They use data from their database about everyone they have contacted to promote their insurance services and try to find the most potential people who can buy insurance. This helps a company to target the most profitable customers and saves time and money for the Insurance Company. Exploratory data analysis performed on the data would help find interesting insights. Predict whether a given customer would like to buy the insurance package, once the corona lockdown ends and travelling resumes. We can be able to help save thousands of rupees of a family..

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

REFERENCES

[1]. Rubi, MaksudaAkter, et al. "Machine Learning Prediction of Consumer Travel Insurance Purchase Behavior." 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2022..

[2]. Rubi, MaksudaAkter, Md Hasan Imam Bijoy, Shanjida Chowdhury, and Md Khairul Islam. "Machine Learning Prediction of Consumer Travel Insurance Purchase Behavior." In 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1-5. IEEE, 2022.

[3]. Rubi MA, Bijoy MH, Chowdhury S, Islam MK. Machine Learning Prediction of Consumer Travel Insurance Purchase Behavior. In2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT) 2022 Oct 3 (pp. 1-5). IEEE.