

A STUDY ON A CAR INSURANCE PURCHASE PREDICTION USING MACHINE LEARNING

Submitted in partial fulfillment of the requirements for the award of Bachelor
of Engineering Degree in Computer Science and Engineering

By

Ravella Varun
REG. NO. 38110453



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119, TAMILNADU**

MARCH 2022

	SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY (DEEMED TO BE UNIVERSITY) Accredited with Grade “A” by NAAC (Established under Section 3 of UGC Act, 1956) JEPPIAAR NAGAR, RAJIV GANDHI SALAI CHENNAI– 600119 www.sathyabama.ac.in	
---	---	---

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **Ravella Varun (Reg. No: 38110453)** who carried out the project entitled “**A STUDY ON A CAR INSURANCE PURCHASE PREDICTION USING MACHINE LEARNING**” under my supervision from November 2021 to March 2022.

Internal Guide

Mrs.Ramya G Franklin.,M.E

Head of the Department

Submitted for Viva voce Examination held on

InternalExaminer

ExternalExaminer

DECLARATION

I, **Ravella Varun** hereby declare that the project report entitled **A STUDY ON A CAR INSURANCE PURCHASE PREDICTION USING MACHINE LEARNING**. done by me under the guidance of **Mrs.Ramya G Franklin.,M.E** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D, Dean**, School of Computing, **Dr. S. Vigneshwari, M.E., Ph.D. and Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department of Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Mrs.Ramya G Franklin.M.E**, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

TRAINING CERTIFICATE

 M-21070010

Certificate of Merit

This is to Certify that RAVELLA VARUN (Reg No. 3810953) has

Successfully completed the internship in MACHINE LEARNING application development in Our

Concern from AUG-10-2021 to AUG-25-2021

During the internship period, the performance of the Internship was found to be GOOD


Program Co-ordinator


CEO

Regd. Office : Green Tree Technologies PVT LTD
No.41, MAS Block, First Floor, Appasamy Street,
West Tambaram, Chennai - 45
Phone : 93840 99077 / 93840 99078 Web : www.gtreechennai.com

Scanned by CamScanner

ABSTRACT

This paper predicted a model that indicates whether to buy a car insurance based on primary health insurance customer data. Currently, automobiles are being used to land transportation and living, and the scope of use and equipment is expanding. This rapid increase in automobiles has caused automobile insurance to emerge as an essential business target for insurance companies. Therefore, if the car insurance sales are predicted and sold using the information of existing health insurance customers, it can generate continuous profits in the insurance company's operating performance. Therefore, this paper aims to analyze existing customer characteristics and implement a predictive model to activate advertisements for customers interested in such auto insurance. The goal of this study is to maximize the profits of insurance companies by devising communication strategies that can optimize business models and profits for customers. This study was conducted through an automobile insurance purchase prediction model was implemented using Health Insurance Cross-sell Prediction data. The proposed system uses Random Forest Classifier. According to the results of this study, Train Accuracy is 0.992 and test Accuracy is 0.936, which has high accuracy. Therefore, the result was that customers with health insurance could induce a positive reaction to auto insurance purchases.

TABLE OF CONTENTS

CHAPTER NO	TOPIC	PAGE NO
1	INTRODUCTION	1
	1.1 PROPOSED ALGORITHMS	1
	1.2 FEATURES OF A RANDOM FOREST ALGORITHM	3
	1.3 CLASSIFICATION IN RANDOM FORESTS	3
	1.4 RANDOM FOREST STEPS	4
	1.5 APPLICATIONS OF RANDOM FOREST	5
	1.6 ADVANTAGES OF RANDOM FOREST	5
	1.7 DISADVANTAGES OF RANDOM FOREST	6
2	LITERATURE SURVEY	7
	2.1 AN ENSEMBLE RANDOM FOREST ALGORITHM FOR INSURANCE BIG DATA ANALYSIS	7
	2.2 INSURANCE CLAIM ANALYSIS USING MACHINE LEARNING ALGORITHMS	8
	2.3 MACHINE LEARNING IN P&C INSURANCE A REVIEW FOR PRICING AND RESERVING	8
	2.4 CANCEL-FOR-ANY-REASON INSURANCE RECOMMENDATION USING CUSTOMER TRANSACTION-BASED CLUSTERING	9
	2.5 A STUDY ON PREDICTION OF BUSINESS STATUS BASED ON MACHINE LEARNING JOURNAL OF KOREAN OF ARTIFICIAL INTELLIGENCE	9
3	METHODOLOGY	11
	3.1 HARDWARE REQUIREMENTS	11
	3.2 SOFTWARE REQUIREMENTS	11
	3.3 LANGUAGE SPECIFICATION	11
	3.4 HISTORY OF PYTHON	11

	3.5. APPLICATION OF PYTHON	12
	3.6 FEATURES OF PYTHON	12
	3.7 FEASIBILITY STUDY	13
	3.7.1 ECONOMICAL FEASIBILITY	13
	3.7.2 TECHNICAL FEASIBILITY	13
	3.7.3 SOCIAL FEASIBILITY	13
4	RESULT AND DISCUSSION	14
5	CONCLUSION	22

CHAPTER 1

INTRODUCTION

- Today, automobiles have become a key tool for land transportation, and their use range is expanding as a means of living.
- Cars are a running weapon and have a problem in that accident caused by banks also increase, leading to a significant increase in life and property damage.
- However, as relationship marketing becomes more critical, the repurchase of insurance and maintenance of insurance contracts by existing customers is an important success factor for the insurance industry.
- In the long term, minimizing the conversion of existing customers to other insurance companies is a positive factor that increases corporate profits.
- Thus, it became a major problem for insurers to increase business efficiency by maintaining existing contractors by controlling the factors of car purchase conversion (Kong et al., 2019).
- However, consumers are concerned that the quality of service of the insurance company they are currently subscribing to is lower than the quality inherently reserved by consumers, and the expected utility from searching for a new insurance company is If it is less than the expected utility, it is highly likely to switch to a sub-optimal insurance company; otherwise, it is highly likely to search for or switch to a new insurance company.
- In particular, the more information about premiums presented by insurance companies, the more likely a subscriber will switch to another insurance company.
- Therefore, this study was conducted by analyzing customer characteristics and implementing a prediction model to activate advertisements for customers interested in such auto insurance in this paper.
- The research results can also be used to increase purchase desire by effectively exposing automobile insurance to existing insurance customers.

1.1 PROPOSED ALGORITHMS

RANDOM FOREST ALGORITHM

Random forest algorithm can use both for classification and the regression kind of problems. In this you are going to learn, how the **random forest algorithm** works in machine learning for the classification task.

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning**, which is a process of combining multiple classifiers to solve a complex problem and to improve the

performance of the model.

A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

The below diagram explains the working of the Random Forest algorithm:

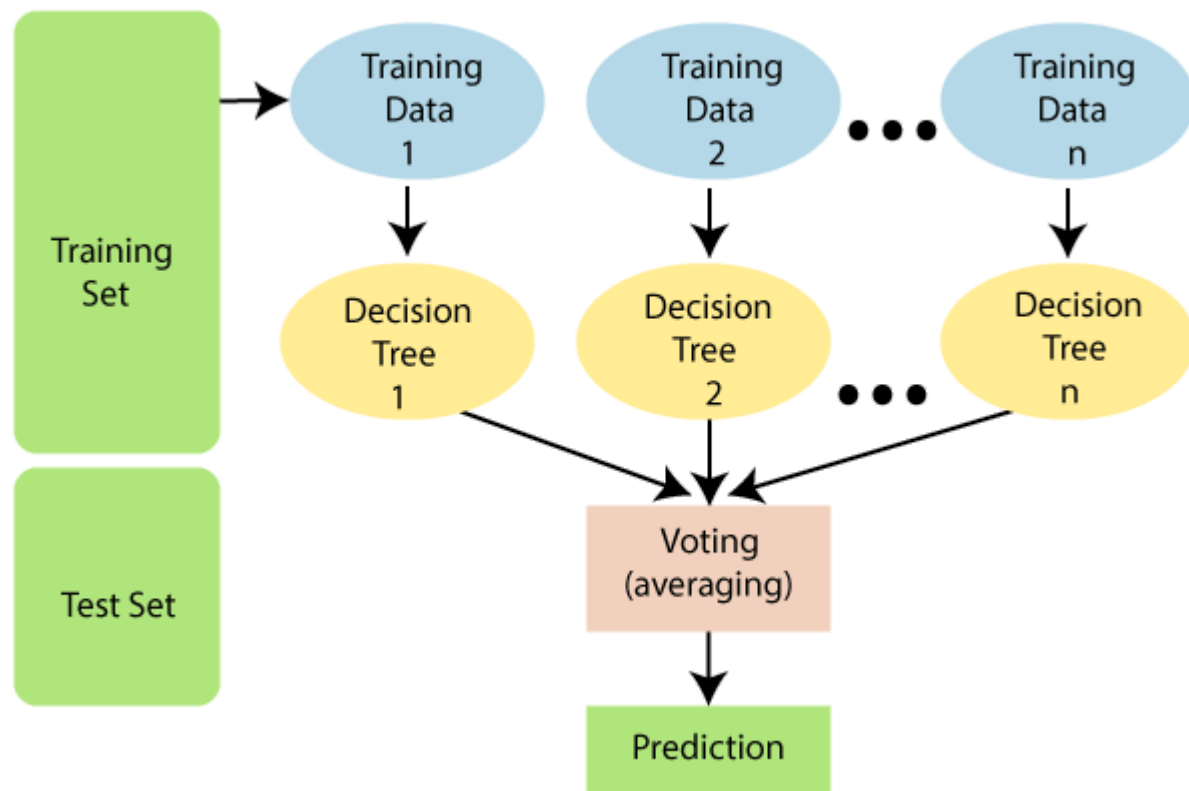


Fig 2: Explaining the working algorithm of the Random Forest algorithm
Below are some points that explain why we should use the Random Forest algorithm:

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.
- It can also maintain accuracy when a large proportion of data is missing.

1.2 FEATURES OF A RANDOM FOREST ALGORITHM

- It's more accurate than the decision tree algorithm.
- It provides an effective way of handling missing data.
- It can produce a reasonable prediction without hyper-parameter tuning.
- It solves the issue of overfitting in decision trees.
- In every random forest tree, a subset of features is selected randomly at the node's splitting point.

1.3 CLASSIFICATION IN RANDOM FORESTS

Classification in random forests employs an ensemble methodology to attain the outcome. The training data is fed to train various decision trees. This dataset consists of observations and features that will be selected randomly during the splitting of nodes.

A rain forest system relies on various decision trees. Every decision tree consists of decision nodes, leaf nodes, and a root node. The leaf node of each tree is the final output produced by that specific decision tree. The selection of the final output follows the majority-voting system. In this case, the output chosen by the majority of the decision trees becomes the final output of the rain forest system. The diagram below shows a simple random forest classifier.

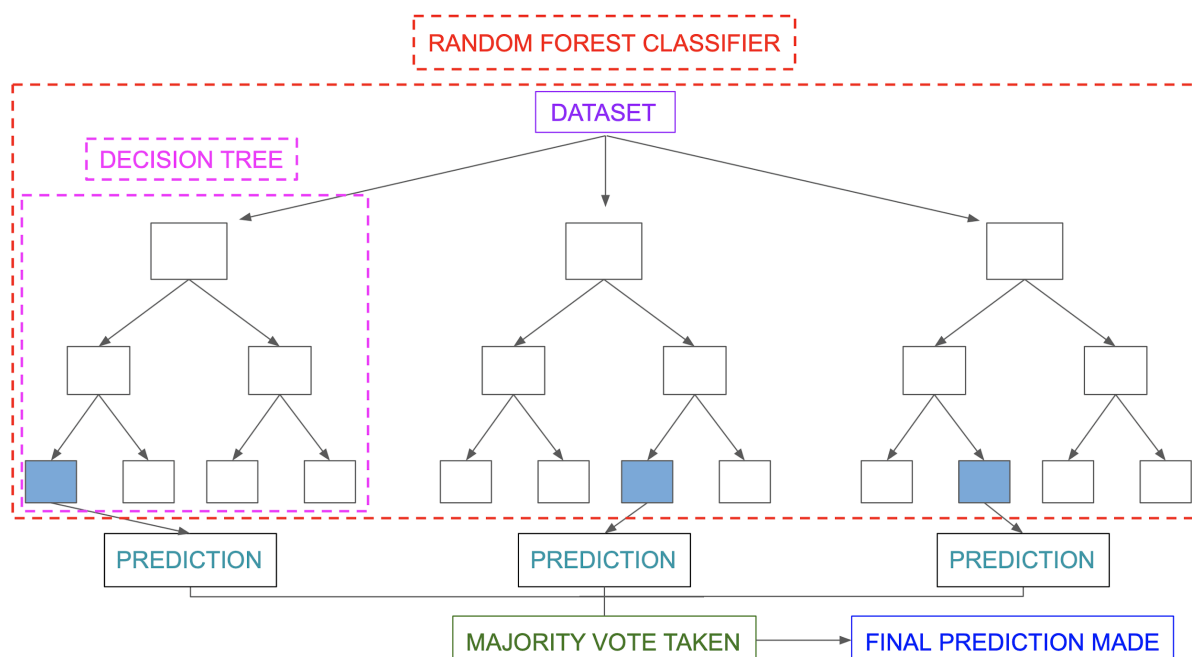


Fig 3: Explaining the Random Forest Classifier

1.4 RANDOM FOREST STEPS

1. Randomly select “k” features from total “m” features.
 1. Where $k \ll m$
2. Among the “k” features, calculate the node “d” using the best split point.
3. Split the node into daughter nodes using the best split.
4. Repeat 1 to 3 steps until “l” number of nodes has been reached.
5. Build forest by repeating steps 1 to 4 for “n” number times to create “n” number of trees.

The beginning of random forest algorithm starts with randomly selecting “k” features out of total “m” features. In the image, you can observe that we are randomly taking features and observations.

The working of the algorithm can be better understood by the below example:

Example: Suppose there is a dataset that contains multiple fruit images. So, this dataset is given to the Random forest classifier. The dataset is divided into subsets and given to each decision tree. During the training phase, each decision tree produces a prediction result, and when a new data point occurs, then based on the majority of results, the Random Forest classifier predicts the final decision. Consider the below image:

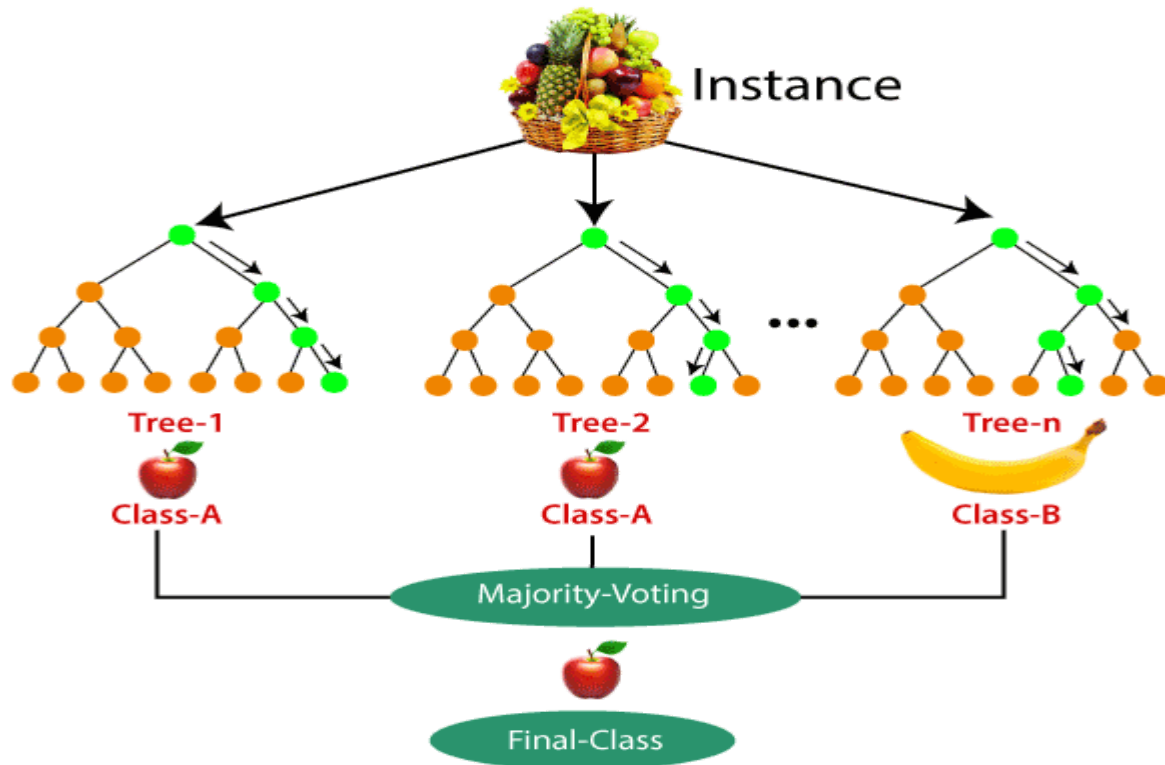


Fig 4: Explaining the Random Forest Classifier algorithm with example

1.5 APPLICATIONS OF RANDOM FOREST

There are mainly four sectors where Random forest mostly used:

1. **Banking:** Banking sector mostly uses this algorithm for the identification of loan risk.
2. **Medicine:** With the help of this algorithm, disease trends and risks of the disease can be identified.
3. **Land Use:** We can identify the areas of similar land use by this algorithm.
4. **Marketing:** Marketing trends can be identified using this algorithm.

1.6 ADVANTAGES OF RANDOM FOREST

Random Forest is capable of performing both Classification and Regression tasks.

- It is capable of handling large datasets with high dimensionality.
- It enhances the accuracy of the model and prevents the overfitting issue.

1.7 DISADVANTAGES OF RANDOM FOREST

- Although random forest can be used for both classification and regression tasks, it is not more suitable for Regression tasks.

CHAPTER 2

LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system. The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

2.1 AN ENSEMBLE RANDOM FOREST ALGORITHM FOR INSURANCE BIG DATA ANALYSIS

Due to the imbalanced distribution of business data, missing user features, and many other reasons, directly using big data techniques on realistic business data tends to deviate from the business goals. It is difficult to model the insurance business data by classification algorithms, such as logistic regression and support vector machine (SVM). In this paper, we exploit a heuristic bootstrap sampling approach combined with the ensemble learning algorithm on the large-scale insurance business data mining, and propose an ensemble random forest algorithm that uses the parallel computing capability and memorycache mechanism optimized by Spark. We collected the insurance business data from China Life Insurance Company to analyze the potential customers using the proposed algorithm. We use F-Measure and G-mean to evaluate the performance of the algorithm. Experiment result shows that the ensemble random forest algorithm outperformed SVM and other classification algorithms in both

performance and accuracy within the imbalanced data, and it is useful for improving the accuracy of product marketing compared to the traditional artificial approach.

2.2 INSURANCE CLAIM ANALYSIS USING MACHINE LEARNING ALGORITHMS

Now a day's Data is playing a central role and is carrying the big asset in the insurance industry. In today's journey insurance industry has a vital role. Insurance transporters have access to more information than ever before. From the past 700+ years in the insurance industry we can consider the three major eras Starting from 15th century to 1960, industry followed the manual era, from 1960s to 2000 we are in the systems era, now we are in digital era i.e. 2001-20X0. The highest corporate object in all three eras is that the fundamental insurance industry has been determined by believing the data analytics in adopting the changing technologies to better and keep the ways and keep capital together. In advanced analysis the main challenge is the analytical models and algorithms which are being insufficient to support insurers; only by machines we can overcome this challenge.

2.3 MACHINE LEARNING IN P&C INSURANCE: A REVIEW FOR PRICING AND RESERVING

In the past 25 years, computer scientists and statisticians developed machine learning algorithms capable of modeling highly nonlinear transformations and interactions of input features. While actuaries use GLMs frequently in practice, only in the past few years have they begun studying these newer algorithms to tackle insurance-related tasks. In this work, we aim to review the applications of machine learning to the actuarial science field and present the current state of the art in ratemaking and reserving. We first give an overview of neural networks, then briefly outline applications of machine learning algorithms in actuarial science tasks. Finally, we summarize the future trends of machine learning for the insurance industry.

2.4 CANCEL-FOR-ANY-REASON INSURANCE RECOMMENDATION USING CUSTOMER TRANSACTION-BASED CLUSTERING

In the travel insurance industry, cancel-for-any-reason insurance, also known as a cancellation protection service (CPS), is a recent attempt to strike a balance between

customer satisfaction and service provider (SP) profits. However, some exceptional circumstances, particularly the COVID-19 pandemic, have led to a dramatic decrease in SP revenues, especially for non-refundable tickets purchased early with CPS. This paper begins by presenting a risk group segmentation of customers in an online ticket reservation system. Then, a CPS fee is recommended depending on the different customer risk groups provided by the cluster segmentation via different clustering algorithms such as centroid-based K-means, hierarchical agglomerative, DBSCAN, and artificial neural network-based SOM algorithms. According to the implemented cluster metrics, which include the Silhouette index, Davies-Bouldin index, Entropy index, and DBCV index, the SOM algorithm presents the most appropriate result. After predicting the new customer cluster, a CPS fee will be calculated with the proposed adaptive CPS method based on the cluster segmentation weights. Determining the weight of each cluster is related to the total CPS revenue threshold for all clusters defined by the SP. Therefore, to avoid a loss for SPs, the total CPS revenue will be kept constant with the threshold that the SP has been adjusted. The experimental results based on real-world data show that the risk group segmentation of customers helps to maintain a balance between CPS fees and SP profits. Finally, according to the calculated weights, the proposed model pegs the SP gain/loss variation with a 0.00012 exchange ratio.

2.5 A STUDY ON PREDICTION OF BUSINESS STATUS BASED ON MACHINE LEARNING. JOURNAL OF KOREAN OF ARTIFICIAL INTELLIGENCE

Korea has a high proportion of self-employment. Many of them start the food business since it does not require high-techs and it is possible to start the business relatively easily compared to many others in business categories. However, the closure rate of the business is also high due to excessive competition and market saturation. Cafés and restaurants are examples of food business where the business analysis is highly important. However, for most of the people who want to start their own business, it is difficult to conduct systematic business analysis such as trade area analysis or to find information for business analysis. Therefore, in this paper, we predicted business status with simple information using Microsoft Azure Machine Learning Studio program. Experimental results showed higher performance than the number of attributes, and it is expected that this artificial intelligence model will be helpful to those who are

self-employed because it can easily predict the business status. The results showed that the overall accuracy was over 60 % and the performance was high compared to the number of attributes. If this model is used, those who prepare for self-employment who are not experts in the business analysis will be able to predict the business status of stores in Seoul with simple attributes.

CHAPTER 3

METHODOLOGY

SYSTEM REQUIREMENTS

3.1 HARDWARE REQUIREMENTS

System	:	Pentium i3 Processor.
Hard Disk	:	500 GB.
Monitor	:	15" LED
Input Devices	:	Keyboard, Mouse
□ Ram	:	4 GB

3.2 SOFTWARE REQUIREMENTS

Operating system	:	Windows 10.
Coding Language	:	Python
Web Framework	:	Flask

3.3 LANGUAGE SPECIFICATION

Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. It was created by Guido van Rossum during 1985- 1990. Like Perl, Python source code is also available under the GNU General Public License (GPL). This **tutorial** gives enough understanding on **Python programming** language.

3.4 HISTORY OF PYTHON

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, Small Talk, and Unix shell and other scripting languages.

Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).

Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

3.5. APPLICATION OF PYTHON

Easy-to-learn – Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.

Easy-to-read – Python code is more clearly defined and visible to the eyes.

Easy-to-maintain – Python's source code is fairly easy-to-maintain.

A broad standard library – Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.

Interactive Mode – Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.

Portable – Python can run on a wide variety of hardware platforms and has the same interface on all platforms.

Extendable – You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.

Databases – Python provides interfaces to all major commercial databases.

GUI Programming – Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.

Scalable – Python provides a better structure and support for large programs than shell scripting.

3.6 FEATURES OF PYTHON

It supports functional and structured programming methods as well as OOP.

It can be used as a scripting language or can be compiled to byte-code for building large applications.

It provides very high-level dynamic data types and supports dynamic type checking.

It supports automatic garbage collection.

It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

3.7 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

The feasibility study investigates the problem and the information needs of the stakeholders. It seeks to determine the resources required to provide an information systems solution, the cost and benefits of such a solution, and the feasibility of such a solution.

The goal of the feasibility study is to consider alternative information systems solutions, evaluate their feasibility, and propose the alternative most suitable to the organization. The feasibility of a proposed solution is evaluated in terms of its components.

3.7.1 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

3.7.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available

technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

3.7.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity.

CHAPTER 4

RESULT AND DISCUSSION SCREENSHOTS



Login

Login

Username

Password

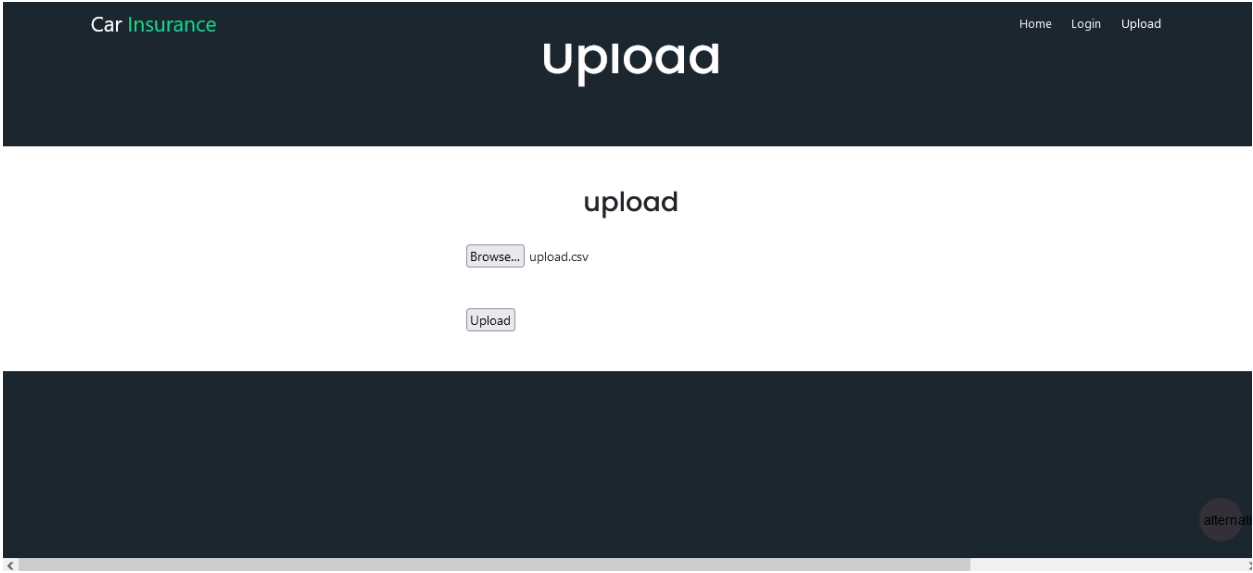
Login

Login

Login

Username

Password



preview

preview

★

	Gender	Age	Driving_License	Region_Code	Previously_Insured	Vehicle_Age	Vehicle_Damage	Annual_Premium	Policy_Sales_Channel	Vintage	Respor
Id											
1	Male	44	1	28	0	> 2 Years	Yes	40454	26	217	1
2	Male	76	1	3	0	1-2 Year	No	33536	26	183	0

preview

★

	Gender	Age	Driving_License	Region_Code	Previously_Insured	Vehicle_Age	Vehicle_Damage	Annual_Premium	Policy_Sales_Channel	Vintage	Respor
Id											
1	Male	44	1	28	0	> 2 Years	Yes	40454	26	217	1
2	Male	76	1	3	0	1-2 Year	No	33536	26	183	0
3	Male	47	1	28	0	> 2 Years	Yes	38294	26	27	1
4	Male	21	1	11	1	< 1 Year	No	28619	152	203	0
5	Female	29	1	41	1	< 1 Year	No	27496	152	39	0
6	Female	24	1	33	0	< 1 Year	Yes	2630	160	176	0
7	Male	23	1	11	0	< 1 Year	Yes	23367	152	249	0
8	Female	56	1	28	0	1-2 Year	Yes	32031	26	72	
9	Female	24	1	3	1	< 1 Year	No	27619	152	28	

alternativ

Car Insurance

Home

Login

Upload

7312	Male	69	1	28	0	1-2 Year	Yes	58079	122	26	0
7313	Male	49	1	8	1	1-2 Year	No	45471	124	99	0
7314	Male	54	1	28	0	1-2 Year	Yes	36139	124	240	0
7315	Male	24	1	45	1	< 1 Year	No	29648	152	35	0
7316	Male	37	1	28	0	1-2 Year	No	30567	23	263	0
7317	Female	22	1	15	0	< 1 Year	Yes	2630	160	160	0

Click to Train | Test



Prediction

car Insurance purchase Prediction

Enter the details

Gender: Female

Age: Age

Driving_License: Customer already has DL

Previously_Insured: Customer already has Ve

Vehicle_Age: > 2 Years

Vehicle_Damage: No

Annual_Premium : Annual_Premium

Vintage : Vintage

Predict

Prediction is :

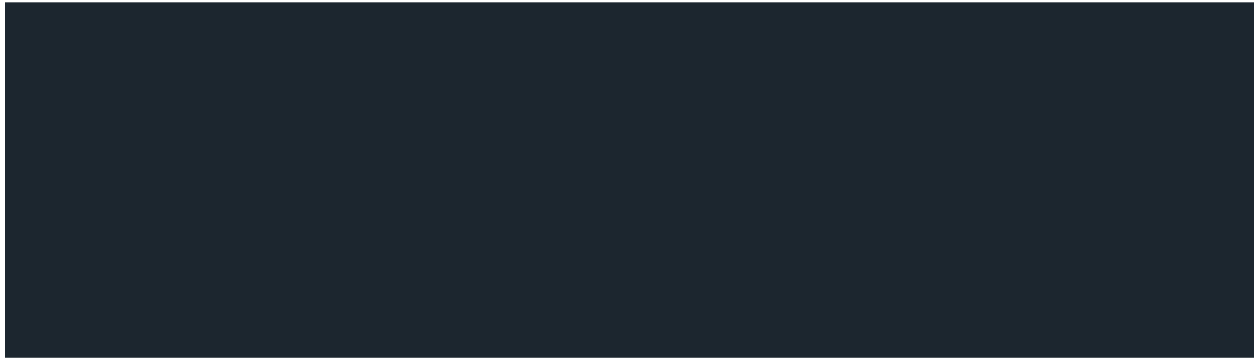
Prediction

car Insurance purchase Prediction

Enter the details

Gender:	Female	
Age:	24	
Driving_License:	Customer already has DL	;
Previously_Insured:	Customer doesn't have V	;;
Vehicle_Age:	Less than 1 Year	
Vehicle_Damage:	Yes	
Annual_Premium :	54408	
Vintage :	86	
Predict		

Prediction is :



CHAPTER 5

CONCLUSION

In this paper, we have discussed that how our proposed system predicts car insurance purchase using machine learning algorithms. The proposed system is also scalable for predicting car insurance on purchase by using techniques after collecting data. The system is not having complex process to predict car insurance that the data like the existing system. Proposed system gives genuine and fast result than existing system. Here in this system we use machine learning algorithms to predict on a car insurance purchase.

REFERENCE

- 1) Choi, B. G. (2018). A study on the meaning of automobile in the no insurance automobile injury insurance. *Journal of Insurance Law Research*, 12(1), 211-234.
- 2) Kong, D. A., & Bang, J. H. (2019). A Study on the Repurchase of Automobile Insurance at Expiration. *Journal of Industrial Economics and Business*, 32(5), 2393-2415.
- 3) Jeon, Y. S. (2014). Analysis on Possibility of Auto Insurance Customer's Company Switching. *Korea Insurance Research Institute*, 25(3), 3-36.
- 4) Kang, M. S., Kang, H. J., Yoo, K. B., Ihm, C. H., & Choi, E. S. (2018). Getting started with Machine Learning using Azure Machine Learning studio. Seoul, Korea: Hanti media.
- 5) Kwak, Y. S. (2020). A Study on Methods to Prevent the Spread of COVID-19 Based on Machine Learning. *Journal of Korean of Artificial Intelligence*, 8(1), 1-6.
Kim, K. P., & Song, S. W. (2018). A Study on Prediction of Business Status Based on Machine Learning. *Journal of Korean of Artificial Intelligence*, 6(2), 23-27.
- 6) You, S. H., & Kang, M. S. (2020). A Study on Methods to Prevent Pima Indians Diabetes using SVM. *Journal of Korean of Artificial Intelligence*, 8(2), 7-10.
- 7) Yoo, S. W., Shin, Y. B., & Shin, D. G. (2020). Comparison of the Machine Learning Models Predicting Lithium-ion Battery Capacity for Remaining Useful Life Estimation. *Journal of The Korean Institute of Gas*, 24(6), 91-97.
- 8) Kim, H. I., Han, K. Y., & Lee, J. Y. (2020). Prediction of Urban Flood Extent by LSTM Model and Logistic Regression. *Journal of The Korean Society of Civil Engineers*, 40(3), 273-283.

- 9) Park, B. J., & Hur, J. (2016). Study on the Analysis of Wind Power Outputs using Multivariate Logistic Regression. Journal of The Korean Institute of Electrical Engineers, 13, 104-105.
- 10) Min, S. H., Oh, Y. J., & Park, S. (2014). User-friendly Robot Vacuum Cleaner System based on Decision Tree. Journal of The Korean Institute of Information Scientists and Engineers, 14, 1836-1838.
- 11) Nam, Y. J., & Shin, W. J. (2019). A Study on Comparison of Lung Cancer Prediction Using Ensemble Machine Learning. Korean Journal of Artificial Intelligence, 7(2), 19-24.

APPENDICES

SOURCE CODE

```
import numpy as np
import pandas as pd
from flask import Flask, request, jsonify, render_template, redirect, flash, send_file
from sklearn.preprocessing import MinMaxScaler
from werkzeug.utils import secure_filename
import pickle
```

```
import numpy as np
import pandas as pd
from flask import Flask, request, jsonify, render_template, redirect, flash, send_file
from sklearn.preprocessing import MinMaxScaler
from werkzeug.utils import secure_filename
import pickle
```

```
import numpy as np
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.preprocessing import StandardScaler
app = Flask(__name__) #Initialize the flask App
```

```
random = pickle.load(open('insurances.pkl','rb'))
@app.route('/')
```

```

@app.route('/index')
def index():
    return render_template('index.html')

@app.route('/chart')
def chart():
    return render_template('chart.html')

#@app.route('/future')
#def future():
#    return render_template('future.html')

@app.route('/login')
def login():
    return render_template('login.html')

@app.route('/upload')
def upload():
    return render_template('upload.html')

@app.route('/preview',methods=["POST"])
def preview():
    if request.method == 'POST':
        dataset = request.files['datasetfile']
        df = pd.read_csv(dataset,encoding = 'unicode_escape')
        df.set_index('Id', inplace=True)
        return render_template("preview.html",df_view = df)

#@app.route('/home')
#def home():
#    return render_template('home.html')

```

```
@app.route('/prediction', methods = ['GET', 'POST'])
def prediction():
    return render_template('prediction.html')
```

```
#@app.route('/upload')
#def upload_file():
#    return render_template('BatchPredict.html')
```

```
@app.route('/predict',methods=['POST'])
def predict():
    int_feature = [x for x in request.form.values()]

    final_features = [np.array(int_feature)]

    result=random.predict(final_features)

    if result == 1:
        result = "Customer is interested"
    else:
        result = 'Customer is not interested'
    return render_template('prediction.html', prediction_text= result)
```

```
@app.route('/performance')
def performance():
    return render_template("")
```

```
if __name__ == "__main__":
```

```
app.run(debug=True)
```