USE OF AI IN CUSTOMER PREDICTION

By Shristi Gupta

ABSTRACT

Artificial Intelligence or AI is a branch of computer science where **machines and computers simulate human intelligence**. In AI, machines are programmed to think like humans and perform tasks that only humans can do. Furthermore, It leverages the problem-solving and decision-making capabilities of human beings.

AI is not just a single technology but a rapidly evolving collection of technologies like deep learning, machine learning, and expert systems. For example, AI systems such as self-driving cars and space-exploring robots learn from the experience. Today there is ongoing research on the applicability of more advanced AI technologies like **Generative Adversarial Networks** and **Edge AI** into various industries.

Currently, AI has been incorporated into many areas of the travel and tourism industry, making lives easier for travelers around the globe.

1. Problem Statement

A travel agency which deals in selling holiday packages provided details of some employees of a company. Among these employees, some opted for the package and some didn't. The company needs to find out in predicting whether an employee will opt for the package or not on the basis of the information given in the data set. Also, the important factors on the basis of which the company will focus on particular employees to sell their packages so that the company can rightly pitch the sales to the interested party only in the process by reducing the advertising cost and increasing the profitability of the company.

2. Market/Customer/Business Need Assessment

We are all in the era of a deadly pandemic where we are pretty uncertain about our future and the pandemic has adversely affected all the business sector and the most affected area is tourism sector and the parties associated with it. So by available dataset we aim to present some useful insights to reduce the marketing cost thereby increase in profit of the company.

3. Target Specifications and Characterization

Throughout the travel search and booking process, there are five key audiences you need to target as purchase intent gradually increases. The audiences we're discussing in this article are:

- **People with no travel plans in mind (yet)**

  Our first target audience doesn't even realise they want to go on holiday yet. They're simply going about their business and your first job is to spark that initial travel interest.

  Social advertising is the ideal place to start with this target audience. This is where people see all their friends' holiday pictures and follow influencers leading the kind of luxury life they dream of.

- **People deciding where to go**
  Once someone decides they're going on a trip, the next key action is choosing where to go. These people aren't casually browsing anymore; they're actively looking for travel destinations and search engines are going to be the first place they turn to.

- **The first bookings (flights, hotels, etc.)**

  With their destination decided, a traveller's next move is to look at flights, hotels, rental cars and any other essentials like travel insurance. Once again, a search engine is the first place people are going to turn to.

- **Secondary bookings (transport, restaurants, attractions, etc.)**

  Once the flights are booked, people turn their attention to planning the best holiday they can in the limited time they've got. During the period between booking flights and departure, travellers have time to do their research; they're going to be looking for the 'unmissable' things to do in their travel destination.

- **People travelling in your area**

  Our final travel target audience is people who are currently in our area, looking for things to see, eat and do. Mobile has empowered travellers to make buying decisions on the move, all they need to do is open up Google Maps to find the best local restaurants, attractions and photo spots.

4. External Search (information sources/references)

We have gathered our information and the dataset from various open source websites.

- https://www.kaggle.com/
- https://link.springer.com/chapter/10.1007/978-3-030-65785-7_7

5. Bench marking alternate products

There are so many advanced AI and machine learning methodologies are in use for correctly predicting and reducing vagueness. We will build the models upon the dataset by implementing LDA and logistic regression methodologies.

- **Linear Discriminant Analysis(LDA)**
  Linear Discriminant Analysis or **Normal Discriminant Analysis** or **Discriminant Function Analysis** is a dimensionality reduction technique that is commonly used for supervised

classification problems. It is used for modelling differences in groups i.e. separating two or more classes. It is used to project the features in higher dimension space into a lower dimension space.
For example, we have two classes and we need to separate them efficiently. Classes can have multiple features. Using only a single feature to classify them may result in some overlapping as shown in the below figure. So, we will keep on increasing the number of features for proper classification.

6. Applicable Regulations (government and environmental)

We are trying to build a model with a obtained dataset by using minimal manpower and systems. We must consider the dataset which should not adversely affect the environment directly or indirectly. i.e. If we should not choose a dataset which is provocative enough to destroy our environment and the data should be complied to the rules and regulations stipulated by government.

7. Applicable Constraints (need for space, budget, and expertise)

- Need For Space – To build the model only a single office room is needed for small scale design implementation.
- Budget-A minimal amount of budget is required for the designing of the models and correctly implementing it.
- Expertise-The most crucial pillar is the man power involved in designing the model and the company must hire an expert with a team of two to do the same.
- Dataset robustness-The top most priority goes to the dataset robustness for successfully implementing any AI to any sector. The dataset must be representing the present market scenario and information gathering must have been done in a good faith to reduce the abnormality of a model and increasing the accuracy.

8. Business Opportunity

During this corona pandemic, day by day new variants being evolved. Like other sectors tourism sector is experiencing huge existential crisis. But as we all know, this tourism sector moulded developing nations into developed ones. So one can imagine the huge potential this tourism sector holds. A small travel company can leverage the market scenario by successfully categorising target customers and spent the resources to convert them as prospective customer. By using artificial intelligence models in the process will be a very cost effective solution.

9. Concept Generation and development

The model will based upon the robustness of the gathered historical data and then we will split the data into train and test. Then we will apply our machine learning algorithm to train data and check train the model for greater accuracy and then test the model for our test set of data and if the model behave normally then we can successfully predict the prospective holiday package buyer from the pool of data .It will greatly reduce operating cost and increase profitability

10. Product details - How does it work? - Data Sources - Algorithms, frameworks, software etc. needed - Team required to develop. - What does it cost? etc

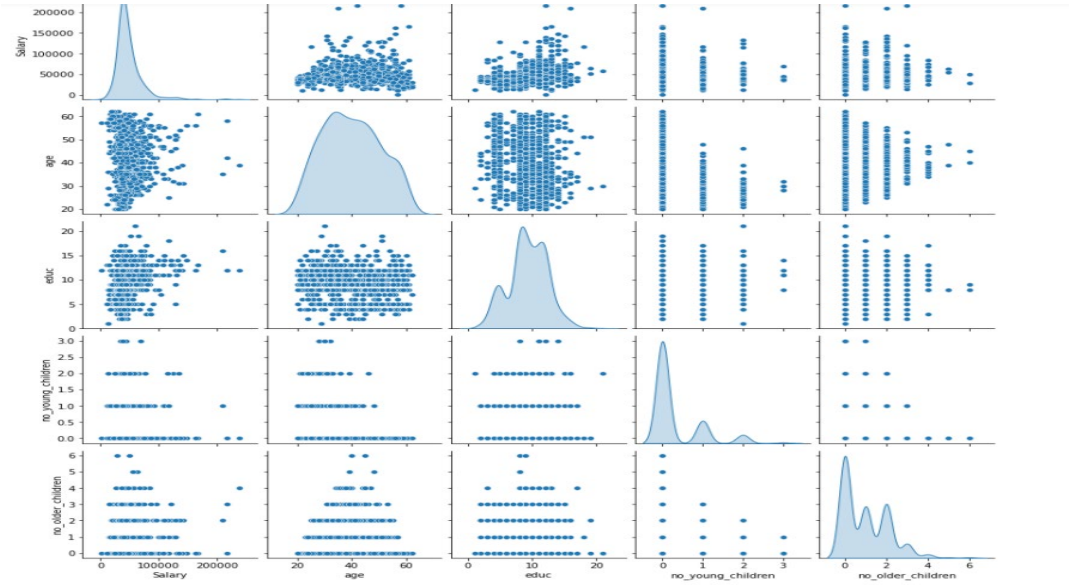
The data set contains the following column and the description is given thereof.

Variable Name	Description
Holiday_Package	Opted for Holiday Package yes/no?
Salary	Employee salary
age	Age in years
edu	Years of formal education
no_young_children	The number of young children (younger than 7 years)
no_older_children	Number of older children
foreign	foreigner Yes/No

Sample of Data set.

	Unnamed: 0	Holliday_Package	Salary	age	educ	no_young_children	no_older_children	foreign
0	1	no	48412	30	8	1	1	no
1	2	yes	37207	45	8	0	1	no
2	3	no	58022	46	9	0	0	no
3	4	no	66503	31	11	2	0	no
4	5	no	66734	44	12	0	2	no

Table 10: Head of dataset



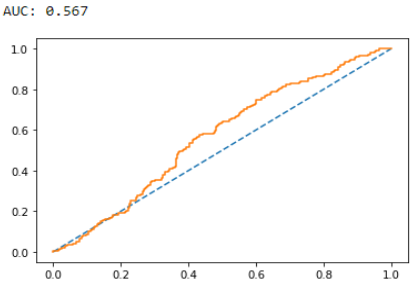
Pairplot

We encoded the data which will help our logistic regression model to perform better.

Logistic Regression:

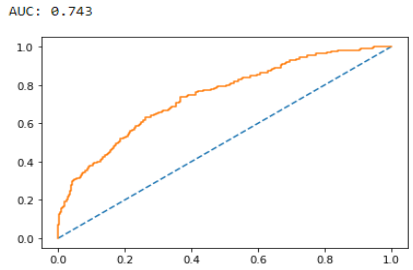
Model 1:

```
0.5305343511450382
[[129  16]
 [107  10]]
              precision    recall  f1-score   support

           0       0.55      0.89      0.68       145
           1       0.38      0.09      0.14       117

    accuracy                           0.53       262
   macro avg       0.47      0.49      0.41       262
weighted avg       0.47      0.53      0.44       262
```



Model 2:

```
0.6374045801526718
[[102  43]
 [ 52  65]]
              precision    recall  f1-score   support

           0       0.66      0.70      0.68       145
           1       0.60      0.56      0.58       117

    accuracy                           0.64       262
   macro avg       0.63      0.63      0.63       262
weighted avg       0.64      0.64      0.64       262
```

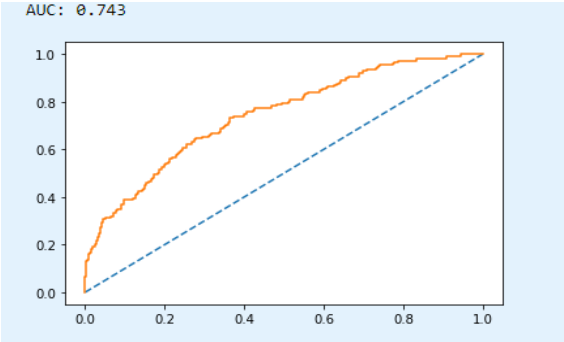


With different parameters we were able to build a second model and improved accuracy of the model.

Model 3:

By using grid search best estimator we again built another model.

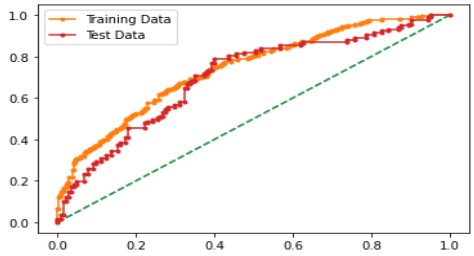
```
0.6450381679389313
[[102  43]
 [ 50  67]]
              precision    recall  f1-score   support

           0       0.67      0.70      0.69       145
           1       0.61      0.57      0.59       117

    accuracy                           0.65       262
   macro avg       0.64      0.64      0.64       262
weighted avg       0.64      0.65      0.64       262
```



Linear Discriminant Analysis:

	precision	recall	f1-score	support
0	0.67	0.75	0.71	471
1	0.66	0.56	0.60	401
accuracy			0.66	872
macro avg	0.66	0.65	0.65	872
weighted avg	0.66	0.66	0.66	872

AUC for the Training Data: 0.742
AUC for the Test Data: 0.703




```
Accuracy is 0.6721311475409836
```

11. Conclusion:

We have successfully implemented different models for the small dataset and successfully narrowed down for prospective package buyer. There are other different algorithms which will increase the accuracy .The profitability of small company can be increased exponentially. From the dataset we could find that people aged above 50 years are not much interested for holiday package. Young people opt for holiday package. Employees age between 30-50 years and salary less than 50,000/- per month opted for holiday packages.

From the dataset we observed that as older people less interested for travel packages, so we need to give discount to them so that we can successfully attract them also.