Travel Insurance Predictions

By Utilizing Data Science Techniques

Presented By Team 7:

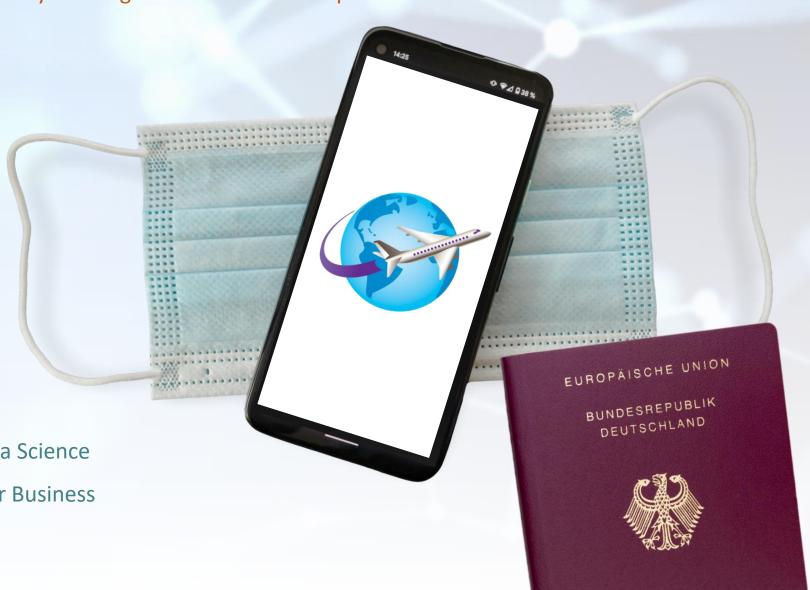
Nima Amin Taghavi Chris Robinson Jimmy Nguyen

Institute: University of San Diego

Degree: Master of Science, Applied Data Science

Class: ADS-505 Applied Data Science for Business

Date: 10/14/2021 (Fall-2021)



Travel Agency

By Utilizing Data Science Techniques

Travel Insurance Predictions

Predict sales of corona virus related travel insurance

Based on a set of commonly collected personal attributes

Analyze the risks associated with making any financial commitments



In Actuality, it became 40% Decline



Not offering pandemic travel insurance



could mean losing sales to another travel agencies which do.

Issues

the agency has no historical sales to make estimates

difficult to calculate the effect not carrying the insurance would have on sales.

predict the likelihood of potential travelers buying COVID insurance

safely cover the annual cost plus 10%

The **total revenue predicted** by the model was ₹ 46,542.00.

The **estimated annual cost** to offer the insurance was ₹ 40,000.00, plus a ₹ 4,000.00 safety margin.

The **total profit** from offering the new insurance is ₹ 2,542.00.

Final Recommendation

Move forward with the covid travel insurance offering

Travel Insurance Predictions

By Utilizing Data Science Techniques

Technical Notebook

2.1 Project Life Cycle

2.2

Data Understanding

Perform Exploratory Analysis to get Insight about the Data and find underlying patterns:

- > Exploring frequency of purchases
- > Look at **Age** distribution and percentage of purchases between Age Groups





- > Effect of **customer's graduation** status on purchasing travel insurance
- > Analyze whether being **Frequent Flyer** or having **Chronic Diseases** effects probability to purchasing travel insurance.

Feature Engineering and Pre-Processing

> Create a feature called wealth groups derived from household income

> Create a feature called Household Size which categorizes customer based on number of Members in the family.

Model and Evaluate Performace

Algorithms picked for Model Prediction:

- 1) Decision tree
- 2) AdaBoost Decision Tree
- 3) Random Forest
- 4) Logistic Regression
- 5) Neural Network
- 6) K-Nearest Neighbors
- 7) Multinomial Naive Bayes
- 8) Linear Discriminant Analysis
- 9) Ensemble Voting Classifier

Splitting the Dataset



- > Class imbalance was not observed
- > Dataset was splitted into 75% Training and 25% Validation.

Hyper-Parameter Tuning

Below are some of the Hyper-Parameter Tuning Performed for each model to maximize F1-Score.

Decision Tree:

min impurity decrease, min samples split

ADA Boost:

Apply Adaptive Boosting with the decision tree as base estimator

Random Forest Model:

max depth, min samples leaf, n estimators

Perform **Feature Importance** technique to decide which columns to include.

Logistic Regression:

stepwise selection by specifying different Penalty Types, Constrain, and Solvers

Neural Network:

hidden layer sizes

K-Nearest Neighbors:

Optimizing K-value by looking at its effect on F1-Score

Evaluate Model Performance

Compare Performance of models using following metrics:

- > Accuracy
- > Precision
- > Recall
- > F1-Measure



Ensemble Learning

Construct an Ensemble **Voting Classifier** using the top 2 models that had the highest F1-Score

(KNN and Random Forest).

Measuring Benefit of All Models

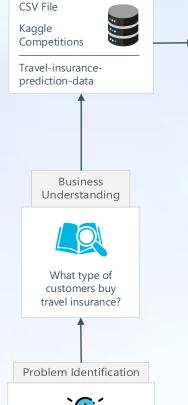
Gains Chart was graphed to aid with finding the optimum model that will make most out of the advertising campaign fundings.

Gather Findings and Give Recommendations

Summarize insights gathered from the data

Give recommendations of how the business can benefit from the findings.





Data Collection



How to utilize Data Science Techniques to Predict potential customers that will buy travel insurance?