

CSCI 531 – Artificial Intelligence Project Proposal

Title:

Airline Ticket Price Prediction

Group Members:

Aneri Moradia – 50227379

Sravani Sagili – 50299631

Lathasree Kelari – 50298602

Problem Statement:

The strategy used by the airline industry to put a price on airfare is derived from the complex structure of various sophisticated rules and numerous mathematical models, which is creating deregulation in this type of industry. The traditional sole factor, which is the distance of how much an airplane travels, is not considered an accurate pricing strategy. Most of the airfare prediction happens either at the national level or regional level (market level). Market-level research is limited.

We can only predict the airfare with the distance traveled which is the market/airport pair between flights origin and destination. To predict using the sole factor, which is distance, is important for all the airline companies to adjust strategy and appoint resources for a specific route.

However, previous research on market-level uses heuristic-based conventional statistical models. This model is known as linear regression which assumes a sequential or linear connection between dependent and independent variables. This has proven wrong in some cases.

In this paper, we address the issue of market section level airfare cost expectation by utilizing freely accessible datasets and a clever AI system to anticipate market fragment level airfare cost.

Data and Data Collection:

Our proposed framework extracts information from two public datasets, “Data_Train” and “Test_Set” – which is a training dataset and testing dataset respectively that consists of all the information regarding airline companies specifically in India which is collected from the Ministry of Civil Aviation Government of India.

Goal / Scope:

The goal is to propose a framework that will draw a comprehensive profile of each market and uses machine learning techniques to predict average airfare on the market segment level.

Modules:

NumPy:

NumPy enriches the programming language Python with powerful data structures, implementing multi-dimensional arrays and matrices. These data structures guarantee efficient calculations with matrices and arrays. The implementation is even aiming at huge matrices and arrays, better known under the heading of "big data". Besides that the module supplies a large library of high-level mathematical functions to operate on these matrices and arrays.

Pandas:

Pandas is a high-level data manipulation tool developed by Wes McKinney. It is built on the NumPy package and its key data structure is called the DataFrames. DataFrames allow you to store and manipulate tabular data in rows of observations and columns of variables.

Sklearn:

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python.

It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use.

Matplotlib:

It is a plotting library used for 2D graphics in python programming language. It can be used in python scripts, shell, web application servers and other graphical user interface toolkits.

There are several toolkits which are available that extend python Matplotlib functionality. Some of them are separate downloads, others can be shipped with the Matplotlib source code but have external dependencies.

System Specifications**Software Requirements:**

OS : Windows

Python IDE : Python 2.7.x and above PyCharm IDE, Anaconda 3.5

Setup tools and pip to be installed for 3.6.x and above

Hardware Requirements:

RAM	: 8GB and Higher
Processor	: Intel i3 and above
Hard Disk	: 500GB Minimum

Expected Result:

Contrasted with the current and late work, our proposed system figures out how to deal with the cost forecast task just by utilizing public information sources with insignificant highlights. Additionally, not confined by a particular market section that normally restricts the current work, this proposed structure can be applied to anticipate the airfare cost for any market.