DEGREE PROJECT PROPOSAL

Carina Wickström
carinawi@kth.se

THESIS TITLE:

Prediction of flight travel demand using machine learning

BACKGROUND:

This is a degree project with the airline company flygresor. They are interested in predicting the flight travel demand for marketing purposes. Hence, they are interested in estimating the impact of different factors that affect flight demand. They have previously used a regression approach in matching events with flight demand swings, to estimate how much certain factors impact the flight demand over time. The factors that have been considered so far are: date, week, holidays, school holidays, version of website, covid outbreaks, warm weather, and snow weather. We want to expand upon this knowledge. We intend to use machine learning to build a predictive model that predicts flight demand. We consider flight demand to be directly linked to the number of flight searches which are logged by the company.

The science of predicting travel demand has been of scientific interest for different applications, and using different techniques [1, 2, 3]. Application areas include targeted marketing [4], as well on a governmental level to support aviation budget planning [5]. Some research has been made on analysing which factors impact flight demand, including national economic factors [6, 7], the impact of AI and robotics [8], marketing [9], and covid [10].

Although the prediction of flight demand has clear scientific importance, taking into account the effect of the fast-varying outside factors previously mentioned (weather, date, week, holidays) and their impact on flight demand is little explored. We want to use these factors and their given effect on flight demand as a base, and use a ML model and additional factors to improve the predictive power.

RESEARCH QUESTION:

We want to improve the predictive power of flight demand per day using ML. With a regression model with the company's provided data as base, we want to add additional features and try different models to improve the accuracy. How does an improved ML model with added features compare to the base model in terms of prediction accuracy?

The base features that we have are: date, week, holidays, school holidays, version of website, covid outbreaks, warm weather, and snow weather. The additional features we are interested in adding are: absolute temperature, temperature in regards to season, global disaster alerts, sport trends, covid news.

RESEARCH METHOD:

I plan to use a regression model to predict the flight demand, using the data provided by the company's model. I will thereafter try to improve the prediction accuracy, using an ARIMA model or ML model. We also want to investigate whether the predictive power increases as we add additional features.

In accordance with answering the research questions, the method will consist of the following steps:

- Calculate a baseline of each factor's impact according to the data provided by the company
- Use the company's provided data in a prediction model to assess the accuracy in predicting flight demand
- Use a more advanced ML model with the same data and find whether the accuracy increases
- Improve the ML prediction model by the following steps:
    - Find more accurate datasets of certain factors (weather, holidays, school holidays, covid outbreaks)
    - Consider additional features not used in the original model (as described in Research Question)
- Observe how much this improved ML model increased the predictive accuracy
- Assess the importance of the different features in the model

BACKGROUND OF THE STUDENT:

My prior work experience and coursework in programming and statistics makes me eligible for a master thesis in this area; I have completed a bachelor in information technology containing various programming courses, Probability Theory and Statistics, Data Mining (project included working with keras), Data-Intensive Computing (project included data analysis using regression), Artificial Intelligence (exchange course where course material included ML and deep learning).

SUPERVISOR AT THE COMPANY/EXTERNAL ORGANIZATION:

Employer: Mattias Nyman (flygresor)
Supervisor: Pierre Echegut (flygresor)

SUGGESTED EXAMINER AT KTH:

Pawel Herman <paherman@kth.se>

SUGGESTED SUPERVISOR AT KTH:

Erik Fransén <erikf@kth.se>

RESOURCES:

The company provides a log of events along with their estimated impact on flight searches. I also received the code of a regression model which was used in producing the above mentioned data. The company further provides me with a log of flight searches, which is the data to be predicted by the output of the ML model.

ELIGIBILITY:

At the time when the degree project starts (p3 2022), I will have completed the sufficient mandatory HP requirement and the mandatory DA2210 Introduction to the Philosophy of Science and Research Methodology for Computer Scientists (ongoing).

STUDY PLANNING:

The courses I have remaining at the start of p3 2022 are:

Machine Learning DD2421 (p3)
Deep Learning in Data Science DD2424 (p4)
Algorithms and complexity DD2352 (divided over p3 and p4)

These courses are left due to exchange studies and a change of master program. I have reason to believe that these three remaining courses match well with my previous knowledge of the topics, and I intend to complete them in parallel with the degree project.

References

[1]
Methods for Travel Pattern Analysis Using Large-Scale Passive Data, LIU, Nils Breyer, Linköping Studies in Science and Technology Dissertation No. 214, 2021, Online: https://www.diva-portal.org/smash/get/diva2:1547806/FULLTEXT01.pdf

[2]
Appraisal of Different Artificial Intelligence Techniques for Travel Demand Analysis, Suchismita Nayak & Dr. Debapratim Pandit, Indian Institute of Science and Technology Kharagpur, Online: http://urbanmobilityindia.in/Upload/Conference/ed85be40-5e4d-44dc-a4a9-ce2d30b78ea7.pdf

[3]
Forecasting airline passengers using designer machine learning - Alexander Backus, Jan van der Vegt, June 26, 2018, Accessed: Dec 15, 2021, Online: https://www.youtube.com/watch?v=A-azJ8qmwug

[4]
Visa Travel Predict (using machine learning), VISA, Online: https://developer.visa.com/capabilities/visa-travel-predict

[5]
FAA Aerospace Forecasts, Federal Aviation Administration, Page last modified: July 06, 2021, Accessed: Dec 15, 2021, Online: https://www.faa.gov/data_research/aviation/aerospace_forecasts/

[6]
Economic factors affecting aviation demand: Practice of EU countries, N. Secilmis and Aylin Koç, January 2016, Online: https://www.researchgate.net/publication/304807370_Economic_factors_affecting_aviation_demand_Practice_of_EU_countries

[7]
External data factors impacting airline profits, Leor Distenfeld, January 17 2020, Accessed: Dec 15, 2021, Online: https://www.meltwater.com/en/blog/the-external-factors-impacting-airline-profits

[8]
Impact of AI and robotics in the tourism sector: a critical insight, Nagaraj Samala et. al, Journal of Tourism Futures ISSN: 2055-5911, April 24 2020, Online: https://www.emerald.com/insight/content/doi/10.1108/JTF-07-2019-0065/full/html

[9]
Factors That Affect The Demand For Air Transportation Tourism Essay, UKEssays, Jan 2015, Accessed: Dec 15, 2021, Online: https://www.ukessays.com/essays/tourism/factors-that-affect-the-demand-for-air-transportation-tourism-essay.php?__cf_chl_captcha_tk__=pmd_DffD7dthhcyu_2GwzMo9yYBxj_CQ95LChyK61.mIH5E-1635355068-0-gqNtZGzNA5CjcnBszQi9

[10]
Changes in air passenger demand as a result of the COVID-19 crisis: using Big Data to inform tourism policy, Inmaculada Gallego and Xavier Font, Journal of Sustainable Tourism, 29:9, 1470-1489, DOI: 10.1080/09669582.2020.1773476, June 05 2020, Online: https://www.tandfonline.com/doi/full/10.1080/09669582.2020.1773476