MIS772 2020 T1

A2 Advanced Predictive Models for Business

AirbnbAI approached you again to develop a RapidMiner process predicting the overall satisfaction of customers renting the New York City AirBnB accommodation, i.e.

- **overall_satisfaction (0..5, including fractions)**

AirbnbAI provided you with an additional sample of 882,000 listings of existing rentals and 803 of listings from Tom Slee's AirBnB repository. The listings include the following information:

- **Room id, name, type, price**
- **Id and name of the property host id and name**
- **Property geo-location, its neighbourhood and borough**
- **The number of reviews**
- **Minimum nights stay**
- **The number of occupants allowed in a rental**
- **Last date and time of posting (date of data collection)**

AirbnbAI would like you to use RapidMiner to generate some insights into the rental listings and these questions are of interests:

A) **Are there any trends in price and customer satisfaction? (initially use visualisation and later with time series analysis)**
B) **What kind of properties receive what type of satisfaction level? (use data clustering and segmentation analysis)**
C) **What is the likely satisfaction level for the new rentals? (by estimation)**

AirbnbAI wants you to use RapidMiner to cleanup and explore the provided data, then develop and evaluate an estimator to predict the customer satisfaction, and to minimise RMSE, MAE and r2.

This assignment aims for students to learn how to ...

- **Articulate problems and solutions in business terms**
- **Gain insights from data**
- **Prepare data for different models**
- **Develop estimation models**
- **Assess and report model performance**
- **Become curious about the world through data and analytics**

DEAKIN BUSINESS SCHOOL

The following mini-case study will be used in assignment A2.

Data: http://www.deakin.edu.au/~jlcybuls/pred/data/Tomslee-AirBnB-NYC.zip
Source: http://tomslee.net/airbnb-data-collection-get-the-data

Individual Tasks and Deliverables

Partial Submission (Question A - marked with the final submission)

Exec Problem: **Define your problem in business terms, in doing so answer question A, cross-reference with other report sections for support.**

Data Preparation: **Deal with duplicates, bad and missing values (use imputation). Transform the selected attributes or create the new ones as needed. Produce supporting charts and tables to answer question A.**

Final Submission (Questions B and C)

Exec Solution: **Describe your solution in business terms, in doing so answer questions B and C, cross-reference with other report sections.**

Data Exploration: **Use clustering and segmentation analysis to investigate satisfaction with rentals. Deal with anomalies. Visualise clusters and anomalies. Provide support for question B.**

Model: **Create three estimation models, i.e. linear regression, decision and an ensemble, to address question C. Explain operators properties.**

Evaluation and Optimisation: **Optimise the model. Evaluate using cross-validation. Include an honest testing. Compare the performance of different models and select the best. Provide support for question C.**

Application: **Apply the best model to new data and investigate results. Also apply a model to a single listing and describe the results.**

See CloudDeakin for more info about this assignment, especially the assignment template and the assessment rubric. Some students will also be asked to present their work to an academic panel, failure to do so will result in zero marks. Late penalty of 5% per day will apply, up to 5 days, at 11:59pm the submission is already 1 day late.