AIRBNB RATING SYSTEM

Predictive Modeling

Andrew Wong

ABOUT ME

Based in Melbourne, Australia. I have worked / lived in 14 countries in Europe and Asia.

A high-performing commercial technologist and solution delivery leader with a track record in working with data science, mobile, digital and software engineering program budget up to AUD$50 million.

I have specialist data science domain in the followings

- ***Agile Development:*** *Rapid iteration of data science projects through small, incremental value delivery.*
- ***Product Development:*** *Developing minimum viable data science / machine learning product (MVP) through Design Thinking and Design Sprinting.*
- ***Industry Verticals:*** *Start-ups, MedTech, EdTech, FinTech*

BUSINESS RESEARCH OBJECTIVES

An Introduction

BUSINESS RESEARCH OBJECTIVES

Rapid Dataset Iteration:
A quick dataset iteration process for data science

Pocket Guidebook:
A friendlier guide to interpreting data science insights

Predictive Modeling:
Engineering features to predict high Airbnb rating

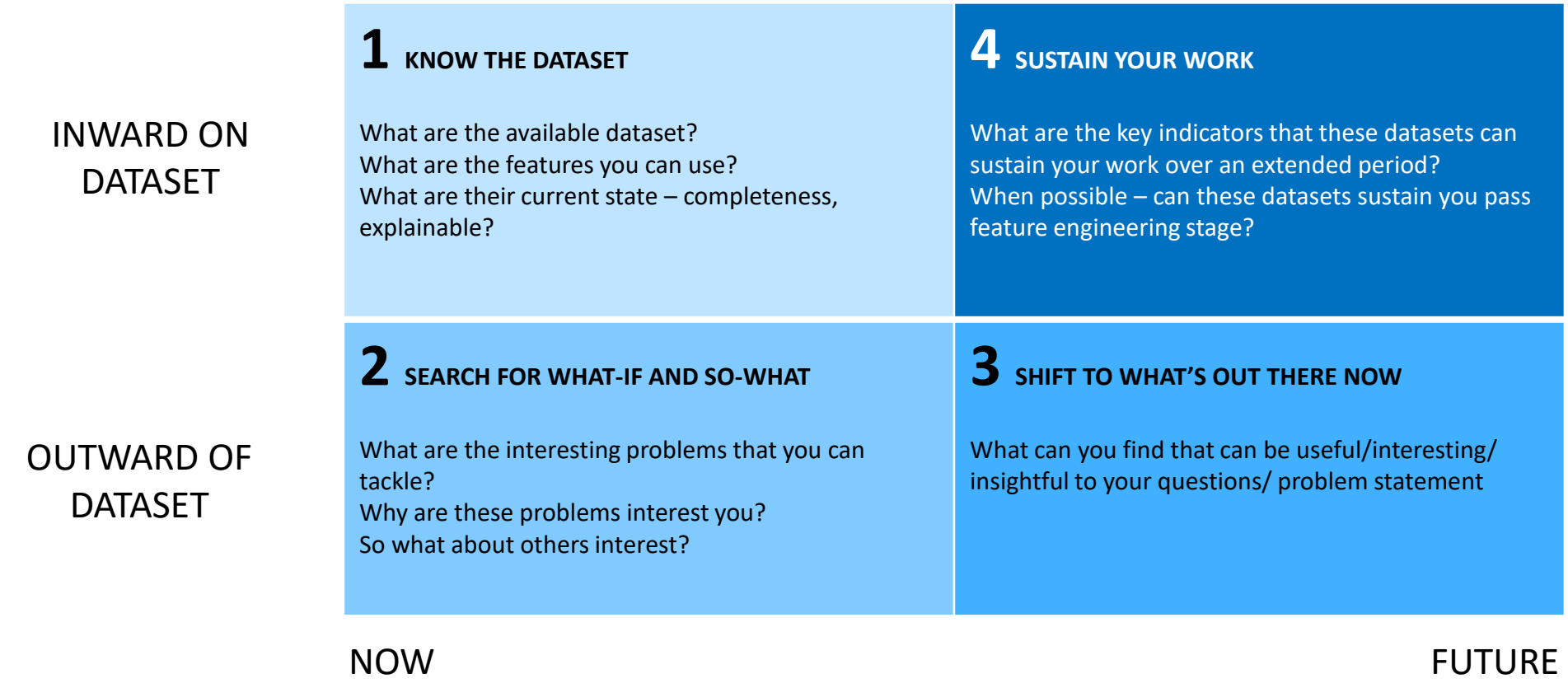
Predictive Modeling:
Determining the influential factors of high rating

BUSINESS RESEARCH OUTCOMES

Data Science +

RAPID DATASET ITERATION

A quick dataset iteration process for data science



POCKET GUIDEBOOK

A friendlier guide to interpreting data science insights

Data Scientist Pocket Guidebook Series

The Quick Wins (and the long game) of Datasets
A guide for aspiring data scientist to iterate datasets rapidly

The Big Three Sticks of Data Science Tasks
A guide for aspiring data scientist to data scrubbing, exploratory data analysis, and feature engineering.

How to Better Design Data Differentiator to Solve Hard Problems
A guide for aspiring data scientist to find the right kind of data

Product Data Scientist Pocket Guidebook Series

Conversational AI in the Age of Hyper Customer Reviews
A forward looking guide for aspiring product data scientist to translate insights into products

Using Data Science to Create Products Customers Want
A guide for aspiring product data scientist create consequential, live changing products

How to Sell your Data Insights to Broader Stakeholders
A guide for aspiring product data scientist to tell a better data story

PREDICTIVE MODELING - 1

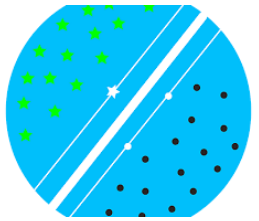
Engineering features to predict high Airbnb rating

Random Forest Model has an accuracy of ~70% on the training set and ~55% on the test set.
This means that we can expect our model to perform with ~55% accuracy on new data.

Random Forest

Training Accuracy : 70.1%
Test Accuracy : 55.5%

SVM

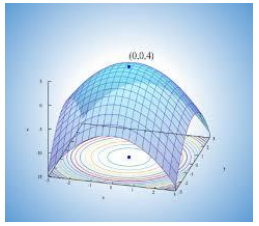


Training Accuracy : 85.8%
Test Accuracy : 49.8%

Logistic Regression

Training Accuracy : 54.0%
Test Accuracy : 54.3%

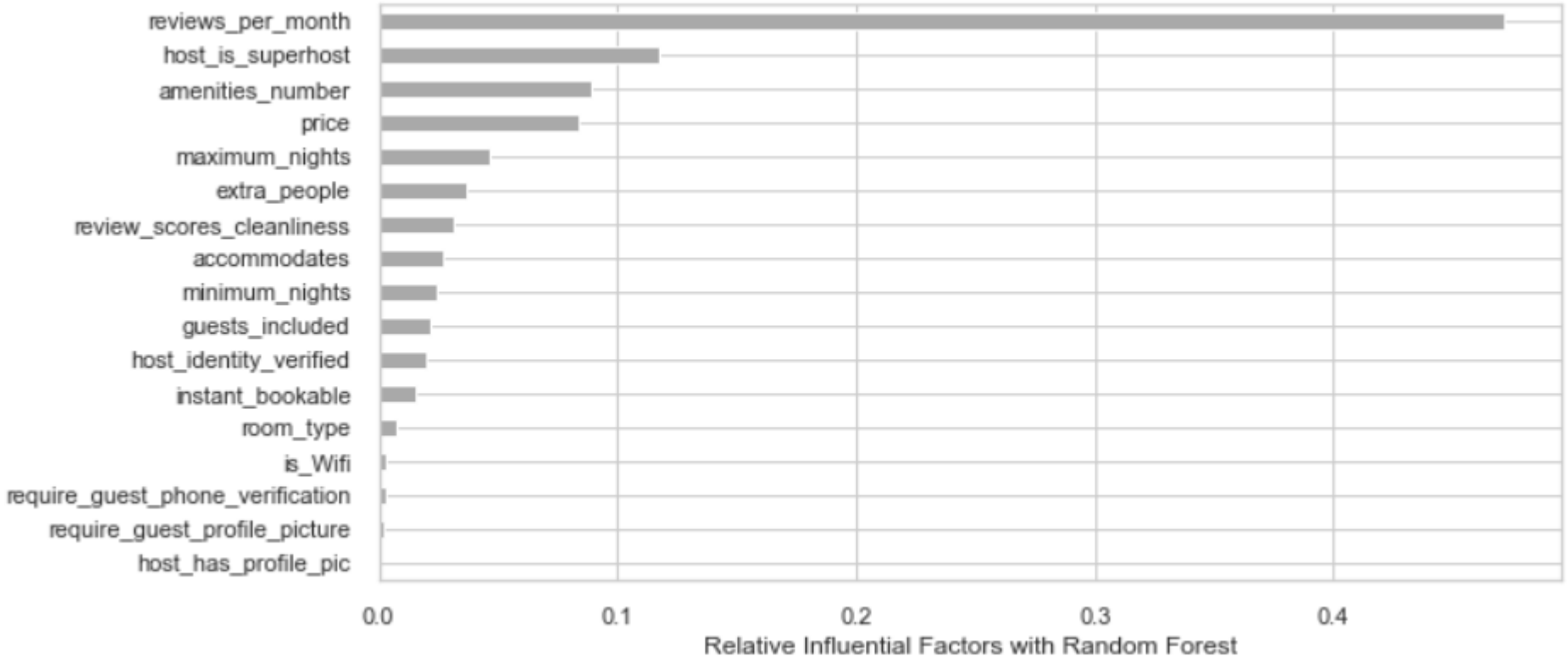
XG Boost



Training Accuracy : 63.2%
Test Accuracy : 55.3%

PREDICTIVE MODELING - 2

Determining the influential factors of high rating

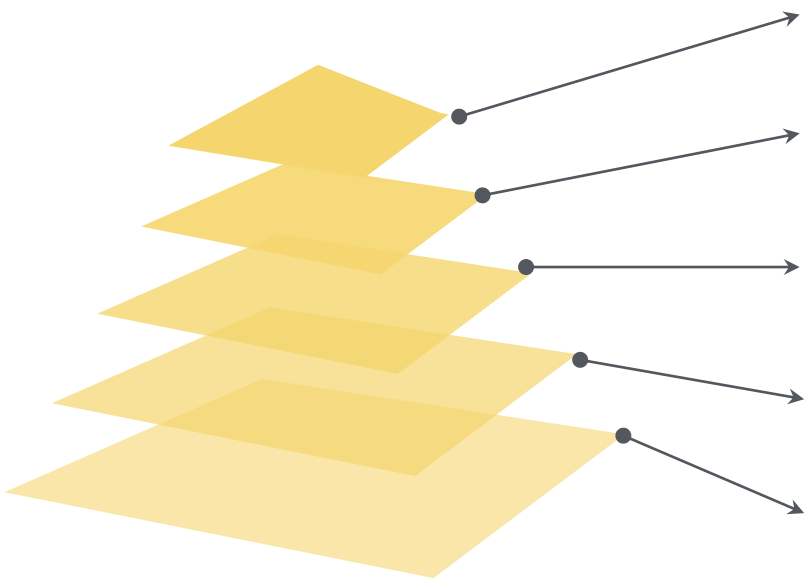


RECOMMENDATIONS

For Business Decision-Makers and Data Scientists

RECOMMENDATIONS FOR BUSINESS DECISION-MAKERS (AIRBNB EXECUTIVE)

How to Better Leverage on Your Data Science Team (DST)



Forward Looking Notes	DST helps to scan the business environmental impact – competitive move, product trends, etc.
Business Vision	DST helps to focus on what are the essential few key drivers of the future (future portfolios; 5-8 years horizon)
Strategic Intent	DST helps to focus on what are the essential few key drivers of the future (current portfolios; 3-5 years horizon)
Yearly Goals	DST helps to drive yearly business insights
Quarterly Goals	DST helps to drive quarterly business insights

RECOMMENDATIONS FOR DATA SCIENTISTS

How to be a Better, Wholesome Data Scientist

Left and Right Side of Data Scientist

LEFT SIDE

Machine, Statistical Centric Data Scientist

- Focusing on means to an end i.e. building an application
- It is not consumable for decision-makers i.e. it requires further interpretation or translation into business-speak.
- It lives in Jupyter Notebook, or production system.
- Metrics: Optimizing, efficiency, accuracy, precision, statistical significant

RIGHT SIDE

Human, Product Centric Data Scientist

- Focusing on the end outcome i.e. drawing insights and decision
- It is about data storytelling
- It lives in executive brief, product management roadmap, design and marketing brief
- Metrics: Revenue, churn rates, customer uplift, what customers love

LIMITATIONS AND FUTURE SCOPE

Next Chapter

LIMITATIONS

The Deep Neural Network (DNN) not working as expected; need to find root cause. *What can do to de-bug?*

Predictive Modeling – limiting to Random Forrest, SVM, XG Boost, and Logistic Regression. *Can I do more?*

FUTURE SCOPE

Re-imagining the future role of Data Science Team

Continue to develop content for the current Pocket Guidebooks

THANK YOU

Andrew Wong

+61 416527928

andrewwongls@outlook.com

LinkedIn: linkedin.com/in/andrewongls

GitHub: github.com/andrewongls

Blog: medium.com/human-science-ai