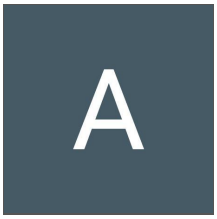
Pradeep Adhokshaja

Data Scientist

Bengaluru | +917259305897 | padhokshaja@gmail.com



Summary

Data Science enthusiast with 3 years of experience in delivering impactful data science projects in the media domain.

Work experience

Data Scientist — 2019 - Present

Equifax Analytics

- **(Research Analytics)Recommendation system prototype that uses users' site behavioral data. The users were classified into clusters using K-Means clustering. For each cluster, we identified the most common articles that were read. Most popular unread articles from each cluster were then recommended to users.**
  - Python (Sci-Kit Learn) for K-Means Clustering
  - Python Pandas for data manipulation
  - Jupyter Notebook for documenting and running each step
  - Google Big Query + Google Analytics for data pull
  - Google Slides to create a Proof of Concept Document for prospective clients
- **Modifying in-house attribution models to account for the absence of user-level digital marketing data for a key channel by coordinating with clients**
  - R (XGBoost) for implementing a Media Mix Modeling -like approach to calculate the importance of missing channel
  - Linux (Cron Job) to automate data pull, model runs and deploying results.
  - Google Slides to create a visual document highlighting the changes from the old approach to the new approach to be used by the client
- **User & Content-Based Recommendation System for one of New Zealand's largest news website to increase user engagement and Average Time spent on site**
  - Tools/ Algorithms used
    - Python & word2vec embedding(Gensim): Used word2vec representations to uncover article to article similarities
    - Python & Google NLP: Used Google NLP to get user personality vectors for frequent users. These vectors were then used to find closest articles that represented the users' tastes
    - Python and sci-kit learn: For the rest of the users, clustering them based on topics read for the week. The clustering algorithm used was K-Means where K was chosen dynamically using a combination of business and mathematical rules. Clustering is done every 24 hours.
    - Python & Numpy: Used a combination of content scoring and behavioral scoring to get a final score between users and articles. Higher the scores, the more relevant the article is to the user's tastes.
    - Google VM instance: Google VM instances were used to run the algorithms with dynamic data and to generate the recommendations, which were then fed to the client's website using API calls.
  - Worked with the Product Engineering team to optimize for speed and scalability.
  - *We were able to increase the average time on site by 15%.*
- **Media Optimization for one of Australia's largest FMCG (with revenue > $AUD1B)**
  - Tools used
    - R (readxl & ggplot2)
    - R(xgboost)
    - Google Sheets for intermediate results

Portfolio & MOOCs

- Data Blog - https://pradeepadhokshaja.wordpress.com/
- MOOCs
  - Machine Learning - Stanford University
- Kaggle Portfolio - https://www.kaggle.com/adhok93

Core Competencies

Business

- Deck Making
- Gathering business insights from data

Machine Learning/Statistics

- **Clustering** - K-Means
- **Linear Models** - Linear Regression , Logistic Regression
- **Non-Linear Models**: XGBoost
- **Inferential Statistics** - ANOVA , Student's t-test , Z-test

Applications

- **Google Slides**
- **Google Docs**
- **Google Data Studio**

Programming languages

- **R** - XGBoost , Caret , GGPLOT2 , Shiny
- **Python** - Sci-Kit Learn , Pandas , Numpy, Jupyter Notebook
- **SQL** - BigQuery

        - Google slides for final results
    - *We provided an optimized channel mix that takes into account previous patterns of expenditure and the COVID-19 scenarios.*
    - *Maximize revenue with a minimal increase (~3%) in overall expenditure. We provided a scenario where Increase in revenue by $AUD 30M was able to be realized, given the data.*

Junior Data Scientist 2017 - 2018

Equifax Analytics

- **Added features based on client requirements to Shiny based dashboards, that were used a prototype to our flagship product "Optimahub".**
    - Shinydashboard
    - R
- **Helped a prominent Telecom Company based in Australia to optimize their social media efforts**
    - Tools used: R (Big Query API and GGPLOT2)
    - *We were able to identify channels & campaigns that worked well with social channels. Our recommendations enabled them to increase social media effectiveness by 20%*
- **End to end delivery of media mix modeling project for a prominent Australian Retail Company. The client was able to decrease significant online and offline spending while keeping revenue constant.**
    - Tools Used:
        - R (XGBoost Gradient Boosting Library to model cost against revenue)
        - R (GGPLOT2 for visualization)
        - Google Slides for Presentation
    - *Media Mix Modeling results and recommendations enabled the client to identify areas of wastage, thereby reducing significant costs while retaining the revenue.*
- **Built a real-time Data Studio reporting dashboard displaying active users and topics across different regions and platforms.**
    - Tools Used
        - Google Data Studio (Dashboard)
        - Used the Google Analytics Real-Time API under R to fetch data in real-time
        - Linux - Using Cron job to ensure automation
    - *This helped to streamline the reporting process of India's largest news aggregator*
- **Consolidated 100+ surveys in Excel format to a presentable format which was then presented through Data Studio Dashboards.**
    - Tools used
        - R (readxl library to conduct ETL of Excel Sheet Data)
        - Data Studio + Big Query to visualize and store data respectively
    - *It helped the marketing team of the client to visualize data much more effectively across geographies.*

Education

Bachelor's of Technology

National Institute of Technology- Karnataka