MRUDUL BANKA

Data Scientist

Internship Experience at Apilation.ai

Data Science Portfolio

LinkedIn Profile

(682) 551-9129

mrudulpopatlal.banka@mavs.uta.edu

mrudulbanka@gmail.com

June 28, 2019

My Internship Experience at Apilation.ai

Contents

About Me:

Hello, I am Mrudul Banka. I like to explore, learn, work and innovate. My areas of interest are - data science, operations research, IoT analytics, robotics, economics, and process improvement - improving customer experience, supply chain, marketing, and advertising.

About my Internship:

- Company Website: http://www.apilation.ai/
- Location: Mill Valley, California - Dallas, Texas - Remote
- Position: Big Data Analytics Co-Op
- Duration: August 2018 - May 2019

About Apilation.ai

Apilation.ai is an intelligent information mining company designed from the ground up to innovate, develop, and deliver on the Big Data analytics needs of our post-IOT data-centric society by providing the people, process, and technologies that empower our clients to 1) Connect ubiquitously, 2) Analyze instinctively, and 3) Act intelligently. All in real-time.

My Internship Experience at Apilation.ai

My Experience:

1. Overall it was a wonderful learning experience which has given me a great outlook of the data science industry i.e.
    - to achieve success as a data science team, the team has to develop a thorough understanding of the business problem along with technical expertise.
    - The team has to work towards developing the trust of decision-makers, dealing patiently with all stakeholders.
    - Another important aspect is how concisely can the team convey the impact which can be achieved by pursuing a project when connecting with the right people.
2. Working on industrial scale Big Data projects showed me how information generated through big data analytics and machine learning can be used to generate and save millions of dollars in revenue.
3. I was provided an opportunity to lead a project where I was the point of contact with the client. This opportunity helped me grow as a professional. I was also asked to train my fellow interns on a technology I was good with.
4. Throughout the internship, I was mentored thoroughly by the CEO which proved helpful to me on a number of occasions.
5. Learned and worked on new programming languages, worked the NoSQL databases with JSonar Studio and MongoDB, how to create visualizations on edge technologies tools. I was able to explore and experiment with all the tools within my sandbox environment which helped me learn faster.
6. This internship prepared me for intense meetings, how to communicate with clients and how to sell your product. My overall communication skills are far superior to what they were before I started the internship.
7. I understood how effective working in an open agile environment is with the key knowledge of how to tackle client requirements and produce deliverables on time.

My Responsibilities:

As described in the offer letter – ''Your job is to help us re-imagine how technology is leveraged in the businesses of the future. You will work under the mentorship of an experienced team leader in either Analytics, Platform, or Experience Engineering; learning new tools, techniques, and skills that can include coding predictive algorithms, building software, and defining how users experience our offerings or all of the above. You can expect to have a substantial influence on the direction that your work, your team, and the company takes as we move forward. Examples of the types of skills we expect for you to possess coming in include, but are not limited to the following: Education in Data Science and Business Analytics and expertise in tools such as SQL, R, Python SAS, etc. Java-Script Notation (JSON), MongoDB, etc. You should understand how queries, joins, and aggregation pipelines work. Education in Web Design and expertise such as the MEAN stack (MongoDB, Express, Angular, Node), HTML5, CSS, etc. SQL skills are also important. Visualization tools such as Tableau, Domo, Spotfire, and PowerBI.''

Along with everything quoted above, I was also made the team lead on the DS-1 reconciliation engine as well as the point of contact for the project.

My Internship Experience at Apilation.ai

Projects Completed at an internship at Apilation.ai

1. To build an automated DS-1 reconciliation engine, analyze the discrepancies, fault detection by rules and its visualization of Tariff # 14 for XXXXX.
2. To develop, analyze and deploy a machine learning model to predict the chances of failure of set-top boxes in the state of XXXX for XXXXX.
3. To build an automated DS-1 reconciliation engine, analyze the discrepancies, fault detection by rules and its visualization of Tariff # 2,3,5,10 for XXXXX.
4. To build a recommendation engine to recommend Tv shows a user would watch given his past viewership and segments to which it belonged.
5. Perform Data Profiling to understand what can be done to obtain business value from their data.

Note: I am unable to post links to my data science projects I had completed during my internship due to proprietary concerns but here is a description of all the work I have been involved in.

Project 1: Automated reconciliation engine – Phase 1

The project is to build an automated DS-1 reconciliation engine. XXXXX has come together through a long series of acquisitions. It is believed that their billing is incorrect in a small percentage of cases, but a small percentage of a lot of circuits can make up a lot of money for the company. For the first phase, they had decided to concentrate just on the single tariff #14 which covers 43% of the 170,000 circuits. The project was divided into two sets of rules based on which we can classify the circuits as billed correctly or billed incorrectly. After classification and implementation of the production version of the application, Processed data, statistical analysis and visualizations were presented to the client. The application was built in over 4 weeks in JSonar studio.

My Contribution to the project:

- Developed logic for exclusion rules and created pipelines in JSonar to execute those rules.
- Created more than 40 pipelines for the production environment of the application.
- Designed statistical parameters through which outputs should be processed.
- Created Visualizations for analysis. Performed analysis and hypothesis testing of incorrectly billed circuits.
- Conducted testing and performed debugging for business rules pipelines developed by other members of the team.
- Presented weekly status update on the project to the client.

Project 2: Machine learning model for Set-top Box

The project is to build a prediction model to detect which set up box is going to fail. XXXXX offers digital television services across the USA. To minimize service order, minimize the time taken to fulfill service orders because of set-top box failures, XXXXX is turning to machine learning to detect which set up box is about to fail based on the data gathered from the set-top box every 15 minutes. For this project, we worked with the dataset gathered from the state of XXXX. An approximate total of 9TB of data was processed with a 90Gb batch of data inflow occurring every 15 mins. 6 different data mining models were created for 6 KPI's which would indicate a fail/no-fail status prediction of the set-top box.

My Internship Experience at Apilation.ai

The application was built in over 4 weeks in JSonar studio, R Studio, and H20 machine learning and python

My Contribution to the project:

- Performed feature engineering to select features, performed principal component analysis.
- Used concepts of regression analysis to find the correlation for time series data.
- Used the H20 machine learning library along with R and python to develop machine learning models for failure prediction. The best model had a prediction accuracy of 91.01% (Model Type: xgboost).
- Tested models developed by other engineers in the team.
- Build production pipelines for a final equation with all 6 KPI's combining to create a score for prediction.

Project 3: Automated reconciliation engine – Phase 2 – Project Lead

The project is to build an automated DS-1 reconciliation engine. XXXXX has come together through a long series of acquisitions. It is believed that their billing is incorrect in a small percentage of cases, but a small percentage of a lot of circuits can make up a lot of money for the company. For the second phase, they had decided to concentrate on multiple tariffs #2, 3, 5, 10, which covers 32% of the 170,000 circuits. The project was divided into two sets of rules based on which we can classify the circuits as billed correctly or billed incorrectly. After classification and implementation of the production version of the application, Processed data, statistical analysis and visualizations were presented to the client. The application was built in over 4 weeks in JSonar studio.

My Contribution to the project:

- I was the project lead for phase 2 and point of contact for the client.
- Developed logic for business rules and created pipelines in JSonar to execute those rules.
- Created more than 85 pipelines for the production environment of the application.
- Designed statistical parameters through which outputs should be processed.
- Created Visualizations for analysis. Performed analysis and hypothesis testing of incorrectly billed circuits.
- Conducted testing and performed debugging for business rules pipelines developed by other members of the team.
- Presented weekly status update on the project to the client.

Project 4: Tv shows recommendation engine

In this project, we were asked to build a recommendation engine to recommend Tv programs a customer would like based on set-top box's viewership history. This recommendation engine could be used as a marketing tool for Set-top users to recommend them shows. A channel churn recommendation engine was also built to understand if a user was likely to churn a particular channel user is currently subscribed to. We used IMDB data to understand what genres the Tv shows belonged to. The application was built in 4 weeks.

My Contribution to the project:

- Building the recommendation engine for Tv shows and channel churn based on cosine similarity, collaborative filtering, and Simon Funk's SVD.

- Performed feature engineering to extract data from JSonar, clean and normalize the necessary fields.
- Create visualizations to take user-id as input and recommend programs
- Created Segments and used the IMDB dataset to enrich the data by adding genres to understand what is being watched.

Project 5: Data Profiling for company XXXX

We received a multinational food company's data. The expectation from us was to look into the data and to understand if any advertising business value can be generated from that data. The data contained a timestamp, location with other parameters relating to the user which cannot be disclosed. My approach was to find answers for 2 questions - 1. How can I help you improve your product? 2. Can we build a new product from the information we possess?

My Contribution to the project:

- Understand the data, definitions for various fields. Cleansed it to provide structure to the data.
- Perform statistical analysis of the data and find if any correlation can be obtained.
- To understand if the ads click rate can be enhanced.
- I came up with two concepts through which their product can be improved with a focus on the location of the user.
- The concepts are being developed with domain expertise from the company.