Priyadharshini Palanikumar

Mobile: +1-682-704-2060 Email: harshinipriya97@gmail.com
Linkedin: https://www.linkedin.com/in/priyapd97 Github: https://github.com/priyapd

Education

University of Texas - Arlington Arlington, TX
Master of Science in Computer Science; Honors in Leadership; GPA-3.42/4 *Aug. 2018 - May 2020*

Coursework: Machine learning, Data modeling and techniques, Big Data, Neural Networks, Database Management System, Human Computer Interaction, Software Engineering, Design and Analysis of Algorithms

Sethu Institute of Technology Virudhunagar, India
B.E in Computer Science and Engineering; GPA-8.3/10 *Aug. 2014 – May. 2018*

Work Experience

Spectrum - Data Scientist Intern May 2019 - Aug 2019, Denver

- Worked closely in **SpecGuide Removal Project**: To predict the customer churn rate from spectrum setup box subscriptions to Internet services and develop customer retention solutions after identifying important metrics
- In the exploratory analysis phase, performed **feature engineering** by **identifying, retrieving, manipulating, cleaning** and **refactoring millions of rows of data** from Hadoop to build new dataset using Hive Query Language
- Created dashboards in **Tableau** for front end Visualization and communicated actionable insights to the Directors
- In the predictive analytics phase, **prototyped machine learning models** such as **logistic regression, random forest and XGBoost**. After several iterations of prototyping, finalized XGBoost model for its best performance to deploy in production. Identified **top 20 Metrics** from 84 features using various feature selection techniques. Expedited the project completion from **20%** to **90%** in 10 weeks which led to the project moving 4 months ahead of schedule

Google Device Lab, Nasscom - Software Developer Intern June 2016, India

- Designed a new product - **'Facebook messenger bot'** that could interact and respond to students about their academic related queries such as assignment deadlines, internal exam marks, class schedules and so on.Each school has an individual customized bot integrated with their own database to fetch relevant answers
- Implemented the conversational bot using **pywit** SDK in wit.ai Interface **(Natural Language Processing Tool)**

Publication

P. Priyadharshini., **Prediction of Diabetes Mellitus using XGBOOST - Gradient Boosting.** National Conference on Recent Innovations in Science, Engineering and Technology (Oct 2017), India.

Technical skills

Programming / DB Lang: Python,R, SQL (Database - MySQL, PostgreSQL, Cassandra)
Libraries: Pandas, Numpy, Scipy, Sklearn, Scikit-learn, Matplotlib, Seaborn, NLTK, Tensorflow
Big data tools / Lang : Hadoop (Mapreduce, HDFS, Hive), Spark (Pyspark, SparkSQL), AWS (RDS,S3,EMR,EC2)
Tools/Framework: Tableau, Dbeaver, Jupyter, Gitlab, RStudio, Apache Aiflow

Relevant Academic Projects

Bachelor's Thesis: Predicting infertility risks in women: Built a Decision Tree Model to predict woman's infertility risks using the dataset collected from various local hospitals and fertility centres. Data exploration, cleaning and modelling were done. Finally, enhanced the accuracy from previously best published accuracy of **74.4%** to **90%**.

Udacity Data engineering Projects: Applied data modeling in Postgresql and Cassandra DB, Built datawarehouse and extracted data from S3 and stored it in RDS through ETL pipeline. Scaled up ETL process by moving datawarehouse to datalake. Automated ETL pipeline and filled up datawarehouse through Apache Airflow and validated the data quality.

Multi-label Classification of abusive comments: Exploratory Data Analysis was performed as the first step. Followed by Data Preprocessing phases which included text cleaning, tokenization and vectorization. In modelling phase, experimented with logistic regression and LSTM to classify abusive comments which resulted in 85% accuracy

Community & Ex-Curricular Experience

Member of LHP, Member of UTA Graduate Council Planning Committee, Fine Arts Secretary - CSE Department, Open Source Contribution - Mozilla Q&A Team (Reporting, triaging, Diagnosing), Co-organizer of two events - GDG