

## SOFTWARE DEVELOPER INTERN

### Experience

Software Developer Intern , 08/2018 to Current

Datawatch â€“ Troy , MI

- Designed and implemented large data pipelines and maintained them using Apache Airflow by building Airflow sensors.
- Designed an end-to-end data driven system that included data ingestion pipelines, data modeling and infrastructure planning built on Spark architecture.
- Migrated large databases for ETL based projects having teams with large data sets using Spark and Presto.
- Developed an automated sentiment analyzing system that streams and stores Tesla tweets on impala instances, detects negative tweets and sends email alerts to the concerned parties.
- The system boosted up the customer feedback analysis process by 40% as compared to manual process.

Data Scientist/Advanced Analytics Intern , 05/2018 to 08/2018

Motion Recruitment â€“ Florham Park , NJ

- Implemented a Topic Modeling Algorithm (LDA using python) for the classification of customer feedback data into various topics of interest/disinterest that helped customer service team to track the major customer issues and quickly act on them.
- Built a Python-powered computational model (using nltk toolkit and NLP tactics) to predict the sentiment of customer feedback that helped identify the areas of improvement to increase consumer satisfaction.
- Worked towards developing and deploying an Android/iOS mobile application using Outsystems framework that connects the employees to the customers that increased customer outreach by 40% and faster redressal of customer grievances.

Data Engineer , 12/2017 to 05/2018

UNIBEES

- Designed a database architecture and developed an automation system to collect, clean and store data for 30,000+ student organizations from Facebook (using Graph API), Twitter (using tweepy), and Instagram with the help of Beautiful Soup, PyMySQL, and SQLAlchemy libraries.
- The data pipeline built helped the team to expand to 100+ organizations and has become the most frequently used application.
- Designed and developed a multi-class classifier for data-tagging on the cleaned data (an improvement upon existing static keyword-based classification) to boost accuracy by 30%.
- Designed and implemented an OCR system using AWS Rekognition, AWS TextInImage to read and tag the images that acted as a sub-system for the automation system reducing manual labor by 90%.
- Projects House price prediction: Developed a system that predicts the selling price of houses by leveraging different features with the help of ML models implemented in Scikit Learn.

### Accomplishments

- MLP Regression, Random Forest (91.3% accuracy) and Linear Regressor, Cross validation, Correlation(using heat maps), forward and backward feature selection Tools: Python(Numpy, Pandas, Scikit learn, standard scalar ,matplotlib, ggplot).
- Online Doctor Appointment: Developed a web app that allows patients to book an appointment with the doctors online and integrates user authentication, authorization and User Access Management.
- Tools: HTML, CSS, PHP, Javascript, JQuery, Mysql Library Management System: A web app for librarians to ease monitoring the transaction of books with salient features like automatic updation of fines on late check-in, text search over book title, genre, author name and ISBN number.
- Automatic Reply Suggestions for ChatBot: An NLP system that suggests a suitable set of replies to a piece of text based on Dialog Act Prediction and Implication Analysis.
- Classification of Babies on the basis of Cuteness- Designed a predictive model based on Deep Neural Networks that determines the measure of cuteness of infants (babies).
- Technology used: Python Tensorflow, neural nets, Java.

### Education

Masters : Computer Science , Present

The University of Texas - City

Computer Science GPA: 3.78 /4.0

Bachelors : Computer Science , May 2017

Indian Institute of Information Technology

Computer Science GPA: 8.39/10 Awarded "Dean's List of Academic Excellence" for 2016 & 2017

### Highlights

- R,Python(Numpy,Pandas, Scikit-learn), Java, C, C++, HTML, CSS, JavaScript, JQuery, PHP, Latex.
- Database : MySQL, MSSQL,Oracle
- Big Data Tools : Hadoop, Spark, Hive, Impala, Apache Airflow, Presto , Impala
- Machine Learning : Categorical Analysis,time series analysis, A/B testing, model selection, dimensional reduction
- Tools : Scikit-learn,AWS-S3,AWS Rekognition api, AWS TextInImage, Graph API, Twitter API, Tableau, Pycharm, Superset, MS Excel, Apache Tomcat, GitHub, WEKA, Eclipse IDE, Liclipse, Matlab.
- Operating Systems : Mac OS, Ubuntu, Windows, Fedora

#### Professional Summary

##### Skills

Apache, API, automation, Big Data, C, C++, CSS, customer service, data modeling, databases, Database, database architecture, Eclipse, email, ETL, features, HTML, PHP, IDE, Java, JavaScript, JQuery, Machine Learning, Mac OS, Matlab, MS Excel, Windows, Modeling, MySQL, NLP, OCR, Operating Systems, Oracle, predict, Python, read, selling, MSSQL, Tableau, Tomcat