

SATHWIK CHENNA

schennam@alummi.cmu.edu | (412) 999-1355 | New York City, NY | [Linkedin.com/in/sathwikcm/](https://www.linkedin.com/in/sathwikcm/)

EDUCATION

Carnegie Mellon University, Pittsburgh

Aug 2017 - Dec 2018

Master of Information Systems Management - Data Science and Analytics. 3.7

WORK EXPERIENCE

Data Scientist, IBM, New York

Feb 2019 – present

Forecasting of Financial Metrics

- Developed time series forecasting models for key business metrics within the finance department. *{time series ML models}*
- Understood the business requirements and did the data engineering accordingly to support all the data analysis and reports.
- Quick prototyping of the data reports, model results and explanation in Python Flask web application.

Employee Attrition Models

- Developed attrition models to predict the likelihood of attrition for associates in the customer-care call centers. *{ML models}*
- Understood the business requirements and did the data engineering accordingly to support all the data analysis and reports.
- Developed aggregated master datasets from flat files, databases and unstructured data for attrition analytics and reports.

Optimal Shift Scheduling Solution

- Developed an optimal scheduling model for the call center associates to maximize the associates' shift preferences and to meet the business call volume demand. *{Linear optimization, Machine learning, data engineering, Python Flask}*
- Understood the business requirements and did the data engineering accordingly to support all the data analysis and reports.
- Quick prototyping of the optimization model and the optimal scheduling results in Python Flask web application.

Cognitive Information Retrieval Solution

- Developed an information retrieval application which interprets the user queries and gets appropriate paragraphs from the PDF files using IBM Watson Compare & Comply and Watson Discovery Service. *{Python, Flask, APIs, data engineering}*
- Understood the business requirements and integrated the IBM Watson services to meet the client needs accordingly.

TECHNICAL SKILLS

Languages: Python, R, Java

Data Science: Machine Learning and Deep Learning algorithms, Statistics, A/B testing, data wrangling and engineering skills

Big Data: Hadoop, Apache Spark, Hive, Kafka, NiFi, ELK, Cassandra and NoSQL databases

Cloud: AWS (EC2, S3, Redshift, Lambda etc.) and Azure

BI and ETL: SQL, data-warehousing skills in Informatica Power Center, Oracle Data Integrator, Oracle BIEE reporting

Others: Scikit-Learn, Keras, Nltk, NumPy, Pandas, Excel skills, Linux, Python Flask, Java web applications.

PREVIOUS WORK EXPERIENCE

Data Scientist Intern, Nomura Securities, New York City

May 2018 - Aug 2018

- Developed supervised models for the classification of financial documents to predict The Federal Banks's opinion on capital markets using Machine Learning and Natural Language Processing and achieved a reasonable accuracy of 87%.
- Worked on Anomaly detection in interest rates for trading opportunities using MLE and Multivariate Gaussian Distribution.

Big Data Engineer, TEKsystems IT Services, Hyderabad, India

Jun 2015 - Jul 2017

Apache Spark-Kafka-Hive Analytics in Hadoop cluster

- Developed the data streaming pipeline NiFi-Kafka-Spark-Hive for real-time analytics with 5-10 secs latency.
- Installed Hortonworks 6-node cluster and wrote data flow procedures using NiFi, Sqoop, Metabase and MySQL.
- Developed Python apps to migrate data into ELK and developed real-time reports in Kibana for log analytics.

Cloudera Hadoop Development – ETL data warehousing

- Installed, managed and handled 3-node Cloudera Hadoop cluster and the admin issues for a team of 15 resources.
- Developed shell and ETL HQL scripts and ran them using Oozie scheduling which brought down the overall ETL data loading process from 36 hours in Informatica to 3.5 hours in Hadoop. Load time reduced by 80%.
- Developed ETL mappings in Informatica 9.x using business logic which decreases ETL load time by 26%.
- Created Facts and Dimensions tables in the Oracle 11g database using Oracle Data Integrator and IKMs and SCDs.

PROJECTS at CMU

Worked on various projects involving data science analysis, ML prediction tasks, classification tasks and full-stack web applications.

Classification of Twitter Spam Tweets: ML classification and text sentiment analysis techniques

Prediction of Public Bus Arrival Time: Multiple Regression techniques in time-series data

Developed a text analytics focused full-stack web application in Java with backend database, middleware layer and UI

Developed an enterprise-level database in Oracle SQL and developed enterprise-level analytics and reports using SQL

Developed common data structures with their applications in Java for data structures and algorithms course.