Sathwik Chenna

cmsathwik7vk@gmail.com | +1 (412) 999-1355 | New York, NY | Linkedin.com/in/sathwikcm

Summary: Data scientist with 5+ years of experience and master's from Carnegie Mellon University

TECHNICAL SKILLS

Languages: Python, SQL
Data Science: Machine Learning algorithms, Regression, Classification, Forecasting, Statistics, A/B Testing
Data Engineering: Hadoop, Spark, Hive, Microservices, Kafka, Kubernetes, Docker, Elastic, NoSQL
Cloud: AWS - EC2, S3, Redshift, Lambda, EMR, RDS, SQS, Step Function, SageMaker, Event Bridge, Load balancer etc.
BI & ETL: Tableau reports and dashboards, Kibana reporting, Informatica, Oracle Data Integrator, Oracle BIEE
Others: Pandas, NumPy, Scikit-Learn, Keras, NLP, linear optimization, hypothesis testing, excel skills

WORK EXPERIENCE

Data Scientist, IBM, New York City March 2019 – present (3.5+ years)

- Worked at the intersection of machine learning, data science, data engineering, and data analytics.
- Took ownership to gather the project requirements, translate them to technical solutions and manage the team resources.
- Developed a cloud data pipeline on AWS using storage gateway, s3, event bridge, lambda, queues, step function, aurora, load balancer, cloud watch and EMR cluster using terraform scripts. {AWS cloud tech}
- Developed PySpark ETL scripts to run data processing jobs on EMR with Livy submissions and step functions.
- Developed end-to-end microservices-based data processing pipelines and applications in an event-driven microservices architecture deployed on Kubernetes cluster. {Python, Kafka, MarkLogic NoSQL db, Oracle db, microservices}
- Developed covid related business KPI Tableau reports and dashboards over large datasets built using PySpark ETL scripts and hive tables {Python, SQL, Tableau, Reports, Hive, Metrics}
- Developed schedule optimization and recommendation model for the call center associates working across different locations and levels to meet the call volume demand and maximize the associates' shift preferences (Python, Flask, Machine Learning, Linear Programming, Data Engineering, Reports, Dashboards)
- Developed attrition prediction models and made manager-level dashboards for call center associates across different locations, departments, and levels. {prediction and classification models, ELK reports, Python}
- Developed prediction models and time-series forecasting models over large datasets to forecast revenue and business metrics for a leading auto insurance company {Python, Time-series models, FB Prophet, model explanation methods}
- Built unified data lakes for ad-hoc analytics from various data sources and databases. {pandas, SQL}
- Developed aggregated master datasets from flat files, databases & unstructured data for scheduling and attrition analytics
- Quick prototyping of data science projects in Flask apps using IBM Watson tech stack and hosted the apps on IBM cloud
- Did agile sprint planning sessions and worked with scrum masters to create JIRA stories, features, bugs, and tasks.
- Managed team resources and relationships with the clients on day-to-day basis and weekly project status check-ins.

Big Data Engineer, TEKsystems IT Services, Hyderabad, India Jun 2015 – Jul 2017 (2 years)

- Developed the data streaming pipeline from network APIs - NiFi - Kafka - Spark Streaming - Hive for real-time analytics
- Developed Python apps to migrate data into ELK and developed real-time reports in Kibana for log analytics
- Developed ETL mappings for facts and dimensions data warehouse tables in Informatica and Oracle Data Integrator

EDUCATION

Carnegie Mellon University, Pittsburgh, PA Aug 2017 - Dec 2018
Master of Information Systems Management, Business Intelligence and Data Analytics

PROJECTS at CMU

Classification of twitter spam vs non-spam: *ML classification and text sentiment analysis techniques*

- Developed statistically significant twitter account-based and tweet text-based features using NLP for the model
- Implemented supervised models with regularization and tuned hyperparameters to develop an ensemble voting classifier
- Studied and analyzed different models with respect to precision, recall, F1, ROC, AUC metric parameters

Prediction of public bus arrival time: *Regression techniques and modeling*

- Worked on timeseries data to predict the arrival time at CMU bus stop using multiple regression and regularization
- Worked on data cleaning, manipulation, engineering part and created features to find an acceptable time estimate

Categorization of android apps: *Text documents classification, NLP*

- Worked on the text descriptions of android applications to categorize them using LDA topic modelling
- Used NLTK package, glove 50d and 100d, word2vec numeric vectors for unsupervised clustering techniques