SATHWIK CHENNA MADHAVUNI

☎ (412) 999-1355 ✉ schennam@andrew.cmu.edu • LinkedIn.com/in/sathwikcm • GitHub.com/sathwikcm

EDUCATION

Carnegie Mellon University, Pittsburgh, USA   Aug 2017 - **December 2018**
Master of Information Systems Management - **Business Intelligence and Data Analytics**   GPA 3.65

Courses: • Practical Data Science   • Machine Learning   • Deep Learning   • Big Data and Distributed Systems
          • Database Management   • Statistics   • Python and Java   • Data Structures and Algorithms

Jawaharlal Nehru Technological University, Hyderabad, India   Aug 2011 – May 2015
Bachelor of Technology - Electronics and Communication Engineering   GPA 3.85

TECHNICAL SKILLS

Languages: C, Java, Python, R, Scala (Apache Spark)
Data Science: Machine Learning algorithms, Deep Learning techniques, Statistics, Natural Language Processing
Big Data: Hadoop, Cassandra, Spark, Hive, Sqoop, Oozie, Kafka, NiFi, Zeppelin, Elastic Search, Cloudera, Hortonworks
BI and ETL: SQL, data-warehousing skills in Informatica Power Center, Oracle Data Integrator, Oracle BIEE
Others: Scikit-Learn, NLTK, Keras, TensorFlow, Numpy, Pandas, Advanced Excel   **OS:** Linux, Windows

WORK EXPERIENCE

Data Science Intern, Nomura Americas Inc, New York City, USA   May 2018 – Aug 2018
Global Markets Research Technology - Data Science (Machine Learning in Python)
- Collected the required data from The Federal Bank's website, in-house trader applications, csv and json files in Python.
- 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.
- Developed Python scripts for Central Bank statements' side-by-side text comparison in PDF using regex and NLP.
- Application Instrumentation using InfluxDB and Grafana: Developed data loading scripts, reports and dashboards.

Big Data Software Engineer, TEKsystems Global IT Services, Hyderabad, India   Jun 2015 – Jul 2017
Apache Spark Analytics in Hadoop cluster   (2 Years)
- Developed NiFi-Kafka-Spark-Hive streaming apps for real-time analytics in Hadoop with 5-10 secs latency.
- Installed Hortonworks 2.4 6-node cluster and wrote data flow procedures using NiFi, Sqoop, Metabase and MySQL.
- Involved in the end-to-end project life cycle: client teleconferences, requirements gathering, knowledge transfer sessions, functional documents, installation, scripts, development and the final data analytics and testing.

Cloudera Hadoop Installation, Maintenance and Support – **Oozie Jobs and Hive Queries**
- 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%.

BI SQL-ETL Development – Informatica and Oracle Data Integrator
- Developed ETL SQL 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.

ACADEMIC PROJECTS at CMU

Classification of Twitter Spam Tweets: ML classification and sentiment analysis techniques   Spring '18
- Collected twitter data from authentic sources for both spam and ham – account information and the tweet object data.
- Developed statistically significant account-based and text-based features which contributed to the spam content.
- Implemented supervised models and tuned hyperparameters to obtain an Ensemble Voting Classifier of 92% accuracy.

Prediction of Pittsburgh Public Transport Bus Arrival Time: Regression techniques   Spring '18
- Worked on timeseries data to predict the arrival time at CMU stop using multiple linear regression and L2 regularization.
- Handled the data cleaning and formatting issues and developed the numerical features to find an acceptable time estimate.

Categorization of Google Play Store Android Apps: Text documents classification   Fall '17
- Worked on the text descriptions of android applications to categorize them using LDA topic modelling.
- Used NLTK, glove vectors, word2vec and different text-to-numbers techniques for unsupervised clustering techniques.

Other Selected Projects: Fall '17 - Spring '18
- **Coursera**: Completed **Stanford Prof. Andrew Ng**'s Deep Learning specialization and Machine Learning courses.
- Implemented a **Collaborative Filtering recommendation** algorithm using popular MovieLens dataset in Python.
- Image classification, object identification and segmentation using **Keras and TensorFlow** and pre-trained models.
- **Advanced Data Analytics in Python**: Yelp Pittsburgh Restaurant data analytics and WhatsApp Group Chat data analytics.
- Implemented **Enterprise Database in Oracle SQL** and developed enterprise-level analytics and reports using SQL.
- Implemented the data structures with their applications in **Java** for **Data structures and Algorithms** course.

Activities: Worked as TA for 95-869 Big Data course. Working as Vice President for CMU Heinz Analytics Club for 2017-18.