VIGNESH THIAGARAJAN

7421 Frankford Rd., Dallas, TX | +1 (469) 571-3764 | vignesh@utdallas.edu | LinkedIn Profile |Github Link

EDUCATION

Master of Science in Computer Science – Data Science Track ***Expected Graduation – May 2021***
The University of Texas at Dallas – GPA 3.83
Coursework: Big Data Management and Analytics, Machine Learning, Artificial Intelligence, Design and Analysis of Algorithms, Statistical Methods for Data Science, Web Programming Languages, Database Design, Operating Systems
Bachelor of Technology in Electrical and Electronics Engineering **Graduated – May 2018**
Graduated First Class with Distinction
Amrita University – GPA 8.64

TECHNICAL SKILLS

Programming Languages	:	Python (Pyspark, Pandas, Numpy, pdb, OpenCV, sklearn), Scala, R, Java
Web Languages	:	HTML5, CSS, Bootstrap, RESTful API, XML, JavaScript, JQuery
Databases	:	MySQL, Oracle DB, NoSQL (MongoDB)
Big Data Technologies	:	HDFS, Apache Spark, Apache Hive, Apache Kafka, Amazon S3, Amazon EMR, Databricks
Tools/IDEs	:	Git, Gerrit, Jira, VS Code, PyCharm, IntelliJ, Visual Studio, Matlab, LabVIEW

WORK EXPERIENCE

Software Engineer Intern - Arm, Austin TX. *JAN 2021 – PRESENT*

- Created **Python** scripts to preprocess data fed to a machine learning algorithm to reduce computational resources used in running verification tools.
- Developed Dashboards for internal use using **Dash Plotly** framework and **MySQL** which provide an overall layout of various system processes and drive technical decision making. Created tests with **Pytest** to test dashboard features.
- Trained ML models on various available data to solve other optimization problems and make analysis.

Data Engineer Intern - ANSYS Inc, San Jose CA. *JAN 2020 – APRIL 2020*

- Helped Automate the processing of large amounts of unstructured data through data pipelines using **Apache Spark (PySpark)**.
- Created Cron jobs to process unstructured data into **Azure Databricks** and create/update Delta tables
- Created dashboards with **PowerBI** and Databricks to help product managers easily consume the data and gain business insights. Documented the workflows and all major changes.

ACADEMIC PROJECTS

Real-time Face Detection and Anonymization (Python, OpenCV) *OCT 2020 – DEC 2020*

- Performed Face detection on stored videos, images and real-time webcam feed with MTCNN and OpenCV.
- Built an Anonymization application of face detection to blur out faces in real-time.

Spark Streaming Twitter Sentiment Analysis (Scala, Spark, Kafka, Kibana, ElasticSearch) *SEP 2019 – DEC 2019*

- Created a Spark streaming application that continuously read data from Twitter about a topic using Apache Kafka. Sentiment Analysis were performed on tweets with a particular topic attached.

Text Summarization (Apache Spark, Scala, Amazon EMR, Amazon S3) *SEP 2019 – DEC 2019*

- Represented sentences in vectorized format and found the similarities between them using cosine similarity.
- Implemented PageRank algorithm to assign ranks to all sentences.
- Classified the sentences as either Summary text or non- Summary text using Decision trees.

Reddit Web Scraper (R, tackscheduleR, ggplot2) *JUL 2019 – AUG 2019*

- Automated a web scraper to periodically scrape the latest news feeds and comments.
- Performed a sentiment analysis on the comments of the latest news and plotted graphs corresponding to the variation of the average sentiment of overall users every hour.

CERTIFICATIONS AND ACHIEVEMENTS

- **Certified LabVIEW Associate Developer**- Received after undergoing **National Instruments'** training program.
- **Published** a technical paper in ***IEEE PEDES Conference, 2018*** : Link