Karneeshwar Sendilkumar Vijaya

+1 (469) 922-7009 | karneeshwarsv@gmail.com | linkedin.com/in/karneeshwar | github.com/karneeshwar

Education

The University of Texas at Dallas — Dallas, TX
Masters of Science in Computer Science — *August 2021 - May 2023*
- Courses: Data Structures and Algorithms, Big-Data Analytics, Artificial Intelligence, Machine Learning

National Institute of Technology, Tiruchirappalli — Trichy, India
Bachelor of Technology — *August 2013 – June 2017*
- Courses: Basics of Programming, Probability & Statistics, Calculus, Operations Research, Rapid Prototyping

Relevant Experience

Copart, Inc — Dallas, TX
Data Scientist Intern — *September 2022 - May 2023*
- Performed Descriptive/Predictive Time-Series Analysis on the data gathered through the different business processes, used the Python Data Science toolkit to access databases, develop pipelines for time-series prediction and draw inferences from the result
- Developed APIs using Flask to perform CRUD operations on a variety of new data based on user requests as an additional feature to an existing vehicle damage score prediction model which has only Create operation
- Built a scheduling system using RabbitMQ, Celery, and Beats that can perform computation every night for the data gathered at the end of a business day eliminating manual processing of data which facilitated other teams using this application in their projects with a 20% reduction in lead time

Renault Nissan Technology Business Centre India (RNTBCI) — Chennai, India
Engineer — *July 2019 - April 2021*
- Developed scripts in Python to automate the post-processing of several simulated crash test results in Metapost, reducing the lead time of projects by 40%
- Saved 1.5M Euros in material and tool cost by developing a linear regression model based on the previous vehicle performance data to predict the design parameters for new vehicle projects on the same platform

Technical Skills

Languages: C, C++, Java, Python, R, PySpark, HTML, CSS, Javascript, MATLAB
Databases: MySQL, MongoDB, HBase, Cassandra
Cloud: AWS, GCP

Projects

Twitter Streaming Sentiment Analysis, UT-Dallas | *PySpark, Kafka, Kibana, Docker* — *July 2022*
- Developed a structured streaming application to perform sentiment analysis on filtered tweets based on keywords
- Twitter API was used to retrieve data and streamed it via Kafka. The tweets were classified as positive or negative using pipeline text classification and data were filtered for the given keywords, for instance: covid19, coronavirus
- Logstash, Elasticsearch, and Kibana were used to store, visualize and analyze the polarity of filtered tweets

Plot Summary Based Search Engine for Movies, UT-Dallas | *PySpark, SparkSQL, Databricks* — *June 2022*
- Built a search engine in Databricks to list top 10 movies which are closely related to the user's search input based on plot summaries using the Carnegie Mellon University's Movie Summary Corpus with over 42000 movie summaries
- The user search input could be either a single term or a query of multiple terms. MapReduce was used to compute TF-IDF and cosine similarity for single-term searches and multi-term queries respectively

DoorDash Database System Design, UT-Dallas | *SQL* — *March 2022 - May 2022*
- Developed an Entity-Relationship diagram to represent all the entities and their relation in the DoorDash system
- Mapped the ER diagram to a Relational Model by following the database normalization rules
- Created tables and implemented 3 triggers and stored procedures using PL/SQL like door-dasher age check, monthly pay stub computation for door-dasher, and overall bill calculation for customer