Hi there!

# I'm Karneeshwar

## Programmer|



Welcome to my personal website. I am a computer science engineer, tech enthusiast, and an avid learner who loves to build cool things.

## Skills

C, C++, Java

Python, HTML5, CSS3

Bootstrap, JavaScript

React, Node.js, Flask

MySQL, MongoDB, Git

Docker, TensorFlow

Hadoop, Apache Spark

Cassandra, R language

MATLAB

## Education

The University of Texas at Dallas

MS in Computer Science

## Work Experience

Copart, Inc

Data Scientist Intern

- Performed Descriptive/Predictive Time-Series Analysis on the data gathered through the different business processes, used Python Data Science toolkit to access database, develop pipelines for time-series prediction and draw inferences on the result
- Developed APIs using Flask to process a variety of new data based on user requests as an additional feature to an existing vehicle damage score prediction model

Renault Nissan Technology Business Center India

Engineer (Automation and Analytics)

- Developed scripts in Python to automate post-processing of several simulated crash test results in Meta-post, reduced 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 in the same platform

## Projects

### Twitter streaming sentiment analysis

- 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 to PySpark. 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

- Built a search engine in Databricks using PySpark 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

### Ranking busiest US airports using Pagerank algorithm

- Developed an algorithm in Databricks using PySpark to rank all the US domestic airports in decreasing order of their business
- The Pagerank algorithm is a famous technique used to find how popular a website is based on the number of hits or how popular a professor is based on the number of citations to their work. Similarly, this algorithm is used here to compute the Pagerank of each airport based on their in- and out-degrees

### Doordash System Design

- 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