

# KALEESHWARAN ARUMUGAM

## Data Scientist and Analytics

+447733851085

kaleesharjun008@gmail.com

<http://www.linkedin.com/in/kaleeshwaran-a-6b603b264>

<https://github.com/Kaleeshwaran08>

Ilford, London

## SUMMARY

At the University of East London, I focused on statistical analysis, machine learning, and data science. I'm skilled in finding patterns and insights to inform strategic choices, having finished projects involving the analysis of huge datasets using Python and SQL. I know how to use Power BI and matplotlib to create graphics that clearly convey findings to stakeholders. My capstone project greatly improved my skills in classification and clustering challenges by applying regression analysis and feature engineering to anticipate house prices. *I developed expertise in developing and carrying out test cases, debugging Python scripts, and guaranteeing software performance and quality while working as a software tester at Data Tracks.* My ability to pay close attention to detail, solve problems, and be proficient in Python were all refined in this role—qualities that are crucial for a career in data science and machine learning.

## EXPERIENCE

11/2020 - 06/2022

Chennai, India

### Software testing And Document Specialist

#### Datatracks

- Developing test strategies and plans to validate software functionality, performance, and security Writing test cases for functions, classes, and methods to ensure they behave as expected.
- utilizing Python testing frameworks such as pytest, unittest, or others to write automated test scripts. Regression testing and pipelines for continuous integration/continuous deployment (CI/CD) depend on automation.
- Conducting manual testing when needed, especially for exploratory testing and usability testing.
- Conducted thorough manual testing to ensure comprehensive coverage of application features.
- Understanding the basics of XBRL and iXBRL, including their purpose and the structure of the documents
- Developed detailed test plans and test cases based on project requirements and specifications.
- Being meticulous in reviewing, editing, and organizing documents is crucial to ensure accuracy and completeness.

## EDUCATION

08/2022 - 08/2023

United Kingdom,  
London

### MSC. Artificial Intelligence

University of East London 2022 -2023  
Data science, Data analytics, Machine Learning, AI

01/2017 - 01/2020

India, Tamil Nadu

### Bachelor of Computer Application

Alagappa university, India 2017 -2020  
Programming, SQL, HTML, Python, Software

## SKILLS AND TOOLS

[Python](#) [MySQL](#) [Data Science](#) [Data Analytics](#) [Data Analysis](#) [Statistics](#) [Machine Learning](#)

[Algorithms](#) [Data Management](#) [NLP](#) [Text Classification](#) [Deep Learning](#) [Web scraping](#)

[Data Visualisation](#) [Pandas](#) [NumPy](#) [PySpark](#) [PowerBi](#) [Tableau](#) [Advanced XL](#) [HTML](#) [CSS](#)

[Django](#) [GitHub](#) [MATLAB](#)

## LANGUAGES

**ENGLISH**  
Proficient



**TAMIL**  
Native



## PROJECTS

### ***Sentiment Analysis of Twitter Data Using NLP***

<https://github.com/Kaleeshwaran08/Dissertation->

- The goal is to develop an advanced sentiment analysis model tailored to Twitter's unique characteristics and evaluate its effectiveness in capturing nuanced sentiment expression.
- The study utilizes the Sentiment140 dataset, consisting of 1.6 million tweets annotated with sentiment polarity labels.
- The research methodology involves systematic data collection, preprocessing, feature extraction, model development, and evaluation. A sentiment analysis model is developed using Machine Learning algorithms such as Naive Bayes, Support Vector Machines (SVM), and Recurrent Neural Networks (RNNs). The model is trained, evaluated, and compared with existing models using metrics like accuracy, precision, recall, and F1 score.

### ***Detecting Social Media Scams Using Deep Learning***

<https://github.com/Kaleeshwaran08/DETECTING-SOCIAL-MEDIA-SCAMS-USING-DEEP-LEARNING>

- This study presents a comprehensive approach to detect and classify spam content on the Twitter platform using deep learning techniques. The research involves preprocessing textual data, including removing URLs, user mentions, hashtags, and special characters.
- The training process involves tokenizing and padding the text data and optimizing the model with binary cross-entropy loss and the Adam optimizer.
- This research provides a valuable tool for enhancing online security and user experience by identifying and mitigating spam content on social media platforms.

### ***Clustering Method and Multi Classifier Incorporating Ensemble Techniques***

[https://github.com/Kaleeshwaran08/Machine\\_Learning\\_on\\_Big\\_data](https://github.com/Kaleeshwaran08/Machine_Learning_on_Big_data)

- In this endeavor Method of Clustering or Recommendation System Using the clustering library with Pyspark to create a KMeans model is dependent on feature selection.
- A short overview of one real time streaming application with the consideration of Legal, Social, and Ethical (LSE) issues.
- And the aim of its unsupervised learning techniques is One recommendation system or clustering methodology, one metaclassifier that uses ensemble techniques, and an overview of the streaming application.
- Enables data-driven decisions to benefit both customers and financial institutions. Provides a comprehensive view of customer financial management and repayment patterns.