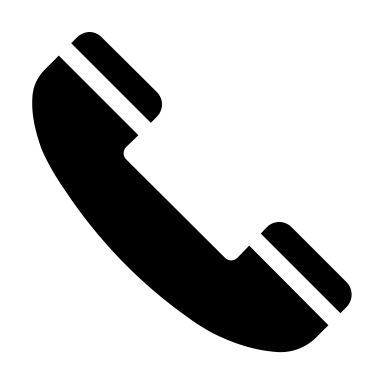
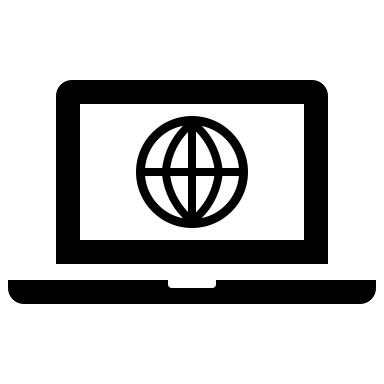
**VIGNESH GIRISH **** - 07760973695

Data Engineer  - [vignesh.malhotra15@gmail.com](mailto:vignesh.malhotra15@gmail.com)

Glasgow, United Kingdom - <https://www.linkedin.com/in/vignesh-malhotra/>



 - [lifeofviggy.co.uk/](https://lifeofviggy.co.uk/)

**WORK EXPERIENCE**

**Data Engineer**

**Adsum Technologies, London 09/2022 – Present**

* Created Data Pipelines on Airflow to ensure complete data flow from source to final prediction models within 10 seconds to 2 minutes depending on the size of the chunk of data.
* Maintaining the integrity of data with constant monitoring, testing and cleaning pipelines setup in each stage of the data flow to ensure high quality and complete data is present for the models with 1 REST API call.
* Training Text classification models on AWS Sagemaker to classify and categorise financial transactions with an accuracy of ~91% and group them accordingly.
* Setting up AWS S3 instances to store and retrieve data from pipelines and update data in PostgreSQL database, ensuring real time data updates for product dashboards within 3 seconds.
* Built a Revenue generation tool using Python, that parses accounting filings data from CompaniesHouse (gov.uk) to identify potential leads for the marketing team and help saving 40-50 hours of market research time per week.
* Defined unit tests for over 250 methods in the codebase performing vital data operations to ensure correct inputs are being passed to the methods and correct outputs are generated.

**ML/Data Engineer**

**CavinKare, India 06/2019 – 08/2021**

* Implemented a deep learning-based face-recognition attendance system to replace traditional biometrics and scaled it up to recognise up to 500 employees.
* Designed and deployed a computer-vision pipeline with a consistent accuracy of around 92% on Microsoft Azure to monitor employee safety and report hazards thus reducing off-time security hours by approximately 30%.
* Developed a parsing tool to generate structured data from unstructured purchase orders and bills using OCR techniques and text similarity saving up to 20 hours of manual data-entry hours per week.
* Analysed sales and logistics data using interactive data visualization to identify seasonality and trends to predict high demand products and peak times and automatically generating supply sheets of over 100 lines of products.
* Setup distributed data pipelines to fetch large volumes of data from Azure Data Lake using Scala and Azure DataBricks within 2-4 seconds.

**SKILLS**

* Tensorflow, PyTorch, NLTK
* Data Analysis (Pandas, Numpy, Matplotlib)
* Docker/Kubernetes/Containers
* CI/CD (Git, Sentry, New Relic)
* Flask/ Django
* Data Visualization (PowerBI, Tableau)
* Python
* ETL, ELT
* SQL (Postgres, MySQL)
* Spark, Hadoop, PySpark
* Azure (DataBricks, ML Studio)
* AWS (S3, Apache Airflow)

**EDUCATION**

**MSc Data Science**

**University of Glasgow, UK09/2021 – 09/2022**

* 2nd Place in Kaggle competition to detect and classify cancer cells

**B. Tech Computer Science**

**SRM University, India07/2015 – 06/2019**

* Graduated First Class (77%)

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

**PROJECTS**

1. **Fetch and parse companies’ filings documents and generate a dataset-** 
   * Send requests to CompaniesHouse API (gov.uk) using Python requests with a company number and type of document to fetch (accounting filing history).
   * Download the linked document and convert each page to an image using Python within 1-2 seconds
   * Extract the required fields from each page using tesseract-ocr.
   * Generate Pandas DataFrame, Clean and format the data – datatypes, missing values, correctness of data and store the final dataset to PostgreSQL database.
2. **Generate music Lyrics after starting words-** 
   * Define bi-directional LSTM model architecture using TensorFlow and Python.
   * Train the model on 12 music artists’ songs to consistent ~87% accuracy.
   * Package the model and the application using Docker and create an image.
   * Deployed the docker image on AWS as a web application which will generate lyrics based on the artist and starting lyrics entered within 5 seconds based on the number of words requested.
3. **Fetch and Evaluate Stock price data of a company selected by the user-** 
   * Send requests to Yahoo Finance API using Python requests with a company ticker and the time period, and load the historical data returned from the Yahoo Finance API as a Pandas DataFrame.
   * Perform Exploratory Data Analysis using Pandas and numpy and plot the Analysed insights using Matplotlib and seaborn.

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

**EXTRA CURRICULAR**

* Runner up in Code for Good Hackathon by JP Morgan Chase & Co. – 2022
* Student Mentor at MCR Pathways, Glasgow – 2022
* Captain of the University Cricket team – SRM University, 2016-2017