HARSHAL HARISH SHAH

+61466341104 | harshalshah71@gmail.com | www.linkedin.com/in/harshalharishshah

OBJECTIVE

Result focused individual and a data enthusiast with a strong understanding of data science and engineering techniques looking forward to applying existing skills in a practical project, and a fast-learning environment. I want to grab the opportunity to grow as an individual and be beneficial to a firm that looks forward to working with me.

SKILLS

Programming Languages	Python, R, BASH, Java, C++, C, SQL, BASH
ML & Visualization Libraries	Scikit-learn, NumPy, Pandas, Keras, NLTK, PyTorch, Matplotlib, Seaborn, Spacy
Miscellaneous Libraries	Dask, MPI, Twitter API, Tweepy, BeautifulSoup, Scrapy, Requests, Jaydebeapi, Paramiko, Office365, SharePoint
Cloud	Basic AWS & Azure, Spartan (University of Melbourne HPC)
Databases / DB Languages	NoSQL (CouchDB), RDBMS, Oracle, MySQL, Postgres
Tools	BitBucket, Git, Microsoft Office, MATLAB, Tableau, PowerBI, Google Data Studio, HubSpot, Jira, DBVisualiser, Putty
Statistics	Descriptive statistics, Distributions, Hypothesis testing, Regression, Variable Elimination, ANOVA
Soft Skills	Communication, Presentations, Documentation, Reporting, Customer Service

EDUCATION

The University of Melbourne, Australia **Feb 2019 – Dec 2020**
Master of Data Science

University of Mumbai, India **Aug 2014 – Jun 2018**
Bachelor of Technology in Computer Engineering

WORK EXPERIENCE

Zetaris Pt Ltd, Melbourne, Australia **Jan 2021 – Present**
Data Scientist | Customer Success Team

- Ingested data from disparate sources using a combination of SQL, Salesforce API's, Redcap API's using python to create a raw data store and further BI processing
- Designed network failure prediction algorithms for a telecommunication client using Machine Learning libraries in python and Zetaris platform
- Developed real time data pipelines for clients and ensured that the data quality is validated & maintained along with the scripts for migrating the data pipelines from one platform to another
- Debug spark errors and improve the efficiency of SQL/Ansi SQL/PostgreSQL queries accordingly
- Maintained pip libraries using JDBC & ODBC driver of Zetaris platform and git repository to ensure code reusability & proper maintenance

Data Disca, Melbourne, Australia **Jun 2020 – Dec 2020**
Intern | Trainee Data Scientist

- Identified the business problems (NLP text classification, different types of visualizations as per the business needs)
- Performed traditional data science approach to solve the given problems (classification, regression, time series) using Python, R, Tableau, and associated libraries

The University of Melbourne, Australia **Feb 2020 – Oct 2020**

Data Science Intern | One CRM Team

- Predicted Student Success using python & various machine learning algorithms from Scikit Learn library
- Attained an overall accuracy of 87% for the prediction system for the RF model after hyperparameters tuning
- Integrated scenario modeling system to the prediction system, to help students who are at the risk of failure

The University of Melbourne, Australia **Dec 2019 – Jan 2020**

Data Science Intern | Campus Analytics Team

- Identified efficiency and sustainability of The University of Melbourne buildings
- Analysed usage of libraries utilizing space and occupancy data (visitors) to understand the overcrowded places
- Achieved an accuracy of approximately 90% for various ML models (SVM, RF) and made it ready for deployment

PROJECTS

Generic Sentiment Analysis Tool for Analyzing Reviews (Python, Flask) **Jun 2020 – Jul 2020**

- Created a generic model to scrape reviews from travel websites to assess reviewer's sentiments and recommend prospective customers accordingly
- The Decision Tree classifier was used to categorize reviews as Positive, Negative and Neutral along with the Text Blob Library and accomplished an overall accuracy of 77%
- Delivered entire project to the client after deploying on the web operating on Python's Flask

Climate Misinformation Detection System (Python) **May 2020 – Jun 2020**

- Worked on the collection of data for climate change information & misinformation leveraging BeautifulSoup and Scrapy libraries in python
- Applied text processing techniques in NLP and generated features from text deploying Vectorisation methods
- Proposed various Machine Learning models such as SVM, Decision Tree, Logistic Regression and Deep Learning models such as RNN, LSTM to restrict texts as 'Climate Change Information' and 'Climate Change Misinformation'
- Succeeded best results (85% percent approx.) using the Logistic Regression model along with TF-IDF, N-gram level vectorised features

City Analytics on the Cloud (Python & Cloud) **Apr 2019 – May 2019**

- Devised a tweet harvesting system by accessing Twitter API and managed to store tweets to CouchDB cluster with the help of Tweepy and CouchDB libraries in python
- Programmed Map-Reduce queries in CouchDB and python to fetch specific data related to the hypothesis

EXTRA CURRICULAR & VOLUNTEER ACTIVITIES

- Working with University of Melbourne students to develop a web scraping & ML based project (Aug 2021 - Present)
- Mentor for Datazen, the official Data Science council at Somaiya Vidyavihar University, India (Jun 2021 - Present)
- Participated in the MYOB Datathon 2020 and Telstra Innovation Hackathon 2020 and developed prototypes (MVP) for the same in September 2020
- Participated & got second prize in the GovHack 2020 Datathon where main aim was to build a prototype of a real time system which detects waste using Deep Learning (CNN) techniques in August 2020
- Alumni Mentor for data science at K. J. Somaiya college of Engineering, India in July 2020
- Peer to peer data science mentor at The University of Melbourne in June 2020
- Implemented a data science based project in MBS Datathon hosted by The University of Melbourne in Aug 2019