Santosh Kumaravel Sundaravadivelu

+61 431933607

santoshisjerry@gmail.com

https://www.linkedin.com/in/santoshkumaravel/

https://github.com/santoshrepo

TECHNICAL SKILLS

Programming Language: Python, R, Java, C/C++

Cloud Technologies: AWS, AWS S3, AWS EC2, AWS Lambda functions, AWS EMR, GCP

Big data technologies: Hadoop, Spark, Maven

Database: MySQL, SQL Server Management Studio, Oracle SQL

Machine learning Tools/libraries: Numpy, Scipy, Scikit-learn, Pandas, NLTK, Gensim, SciPy, Dask, Tweepy

Deep Learning Frameworks: Tensorflow, Keras, PyTorch, Word2Vec, Doc2Vec

Pre-trained Deep learning models: Google's BERT

Data Visualization Tools: Tableau, PowerBI, Google Charts

Data visualization libraries: Matplotlib, Seaborn, ggplot, Plotly, geoplotlib, leaflets, R shiny, shiny dashboards, dash

Application Server: JBoss, Websphere

EMPLOYMENT HISTORY

Voop.Global February2020 – June2020
Data Scientist Intern Melbourne, Victoria

- Mined the data from different sources such as twitter, Google trends, RSS Feeds, Wikipedia using various libraries and API like **Tweepy API, Wikipedia API**; built the knowledge base with the mined data using **Python**.
- Using the **RSS Feeds**, Covid19 Voop Dictionaries was created. For instance, Unemployment, Social connection are examples of Covid19 Dictionary.
- In the GCP **environment**, the **BERT** Deep learning model was trained on the mined data to determine the background.
- Dot product and Cosine similarities were used to segregate the context(mined data) based on the Covid19 Voop Dictionary.
- The dashboard was created in **PowerBI** for visualization(https://voop.global/covid19-analysis).

RMIT University September2019 – March2020
Research Assistant Melbourne, Victoria

- Used **pdfminer** library in **Python** to extract the contents in the PDF.
- **Topic modeling** was performed to understand the topics discussed by the local council in Victoria.
- **Sentimental analysis** was done over the topics to segregate the positive and negative sentiments.
- **Doc2Vec Algorithm** was built to understand the document from the local councils.
- **Cosine similarities** are calculated for the degree of sustainability among the councils.

Newgen Software Technologies June2017 - June2018
Software Engineer Chennai, India

- Identified patterns in the data and translated them into insights on integration issues and vulnerabilities with improved solutions using **Agile Methodology**.
- Setting up the **dev environment** and **deployment** in the production environment using **Java** and **Python**.
- Interacted with the **Stakeholders** for better implementation and weekly progress.
- Versed in the complete Software life cycle from preliminary needs analysis to enterprise-wide deployment and support.
- Extensively worked on data extraction, loading data, Transformation, from various sources like **Oracle** and **SQL Server**.

LTI - Larsen & Toubro Infotech December2016 – February2017
Software Engineering Intern Chennai, India

- Implemented graphical interface to manage VDIs(BOTS), which makes it much easier than the existing system
- Developed the interface using **HTML, CSS, Javascript**, and deployed in the **dev** environment.
- Analyzed insights and reported, and documented.

EDUCATION

Master of Data Science, RMIT University Melbourne, Australia
Grade: 75% July 2018 – July 2020

Bachelor's in Information Technology, Sri Sai Ram Engineering College Chennai, India
Grade: 74% June 2013 - June 2017

PROJECTS

Types of accidents in Victoria Neighbourhood February 2020 – June 2020

- The application was built in **R shiny**.
- **HTML** was also integrated with the shiny app to build a better visualization.
- **Leaflets** library was used to visualize the maps and find the correlation between the accidents and house types.
- Different types of graphs were created to contribute to the analysis of the accidents. For instance, a Stacked bar chart was plotted to understand the number of accidents in the Suburbs.

Analyze the egg Depositions of the Lake Huron Bloasters February2020 – June2020

- Converted the data frame into **time series** data using **R**.
- **SARIMA** models were built to find best model specification **ACF/PACF** plots **BIC** and **EACF** plots were plotted and analyzed.
- Estimated the model parameters and significance tests using **LSE** and/or **MLE**.
- The best fit model was selected and **forecasted** the values for the upcoming 5 years.

Performance Comparison of Map-Reduce Algorithm July2019 – November2019

- Analyzed two big datasets Common Crawl and Amazon Review from **AWS S3** bucket.
- Established a connection with the jump host and accessed the **AWS EMR** Master.
- Analyzed the Performance with boosting the clusters, changing the data sizes among the datasets, changing the number of mappers and reducers.

Analyzing and Tracking the sentiment and topics on Social Media July 2019 – November 2019

- Connected with twitter was established by **Tweepy API** and fetched around 10k tweets over one week.
- The tweets were then cleaned and then normalized and tokenized to find the sentiment of the words/sentences.
- The trend of the Oneplus was found by the hashtags which were visualized to find the interesting findings.
- Implemented **Topic modeling to** find the relevant topics discussed in the tweets.
- Performed **sentiment analysis** on the tweets using the **Vader** sentiment.

Predicting the revenue decline for Portuguese Banking institution(Kaggle Project) February2019 – June2019

- Using **Jupyter Notebook** the Portuguese Banking dataset was loaded from Kaggle.
- Data Exploration was done in **Tableau** to find interesting insights
- **KNN** and **Random Forest Machine learning models** were built to predict the decline in the revenue in the institution.

AWARDS AND HONOUR

Best Performer Award for the Security Bank, Newgen Software Chennai, India
February 2018

Pat on the Back Award, Newgen Software
Chennai, India
May 2018

EXTRA-CURRICULAR ACTIVITIES

Student Ambassador, RMIT University Melbourne, Australia
June 2019 – Present

Event Volunteer, RMIT Student Union Melbourne, Australia
October 2019

CERTIFICATES

Data Literacy. RMIT University Melbourne, Australia
April 2020

Academic Integrity Awareness, RMIT University Melbourne, Australia
April 2020