Sai Vineeth Kandappareddigari

Boston MA| (857) 424 7512| kandappareddigari.s@husky.neu.edu| Linkedin|Github | Orcid

EDUCATION

Northeastern University, Boston, MA

Master of Science, Data Analytics Engineering, **GPA 4.0/4.0** Expected May 2021

Courses: Data Mining, Probability & Statistics, Adv Machine Learning, Deep Learning, NLP, Computation & Visualization, DWBI

Manipal Institute of Technology, Manipal, India

Bachelor of Technology, Computer and Communication, **GPA 7.93/10.00 (with distinction)** May 2017

WORK EXPERIENCE

Research Assistant, Northeastern University Aug 2020 - Present

- Building Deep Learning Models for Human-Machine Interaction Lab related to Health Care for Pain Recognition leveraging CNN (VGG16, RESNET, Inception) models and designing the **AWS (Kinesis, lambda, Sagemaker) pipeline.**

Honeywell Aug 2017 - Jul 2019

Data Scientist

- Developed and monitored **predictive models (Supervised**) for the refinery plants and supported in the decision making with SMEs; worked on classification models (**LDA, Logistic, SVM** and **decision trees**)
- Designed and implemented data pipelines to stream sensor data on memSQL (in-memory DDB) using **pySpark, Kafka and ETL scripts(python)**; Resulting in reducing load and latency by 60%; Analyzed massive sensor timestamp data using **Influxdb**
- Developed, analyzed, and resolved issues of a group of 5 MVC web applications in **C#.NET** utilizes **ML.Net** in an **agile** team, and formulated **queries**, **stored procedures**, **pivot tables**, and **transactions** using **SQL** for **data management**, with an added recommendation system used by internal customers of refinery plants
- Reduced delay in addressing client issues, by developing Distribution Mail List Bot and Knowledge Base with a retention rate **~36%**, to address common issues, using internal **LUIS**, **NLP**, **MongoDB**
- Elected as focal leader for Innovation activities within the department and received '**Best Kaizen**' and '**Honeywell Star**' awards

Honeywell Jan 2017 - Jul 2017

Research Intern

- Analyzed difficulties faced by employees in refinery plants and developed an internal bot service which helps in finding documents that are required, using **NLTK** , **Scikit**, **Tensorflow** and **MongoDB**
- Reduced the categorical features with the added **weights** for feature reduction using **Theil's U** that minimized model processing time by **30%** for designing refinery plants

PROJECTS

Abstractive and Extractive Research Paper Text Summarization

- Developed a **seq2seq Bidirectional LSTM Encoder** with an **attention** layer and unidirectional decoder using **Glove** embeddings model for summarization and to optimize time spent on research papers by academics during their research work
- Technologies Used: **Python, Keras, Tensorflow, Glove, DeepNeuralNet, Transformers, API, AWS**

Time Series Forecasting of recommended stocks for the users of same cohort

- Applied **EDA**, **feature Engineering** to the crawled stock recommended data and trained it on **ARIMA, Fourier, LSTM** with **hyperparameter tuning** , where stock price movement is analyzed and close prices are predicted with an accuracy **~**68.9%
- Technologies Used: **Tensorflow, Keras, Scikit, Web Crawler, Sequential and Forecasting Models, Hyperparameter Tuning**

Demographic Customer Segmentation and Acquisition based on Financial Company's Customer data

- Clustered customers into segments based on demographic data using unsupervised learning (Kmeans, DBScan, PCA) & predicted potential customers using classification models(LDA, Logistic ,RF, XGboost), with XGboost achieving a ROC score of 0.82
- Technologies Used: **Feature Engineering, Dimensionality Reduction, Clustering, Classification Models, Hyperparameter Tuning**

Demographic Trend analysis on Suicide rate and it's relation with Happiness

- Analyzed(EDA) presented high-level interactive visualizations of Suicide rate & examined the "Suicide in Happy places" paradox
- Technologies Used: **Anomaly detection, Feature Engineering, Data wrangling, Data Transformation, R (gganimate), Tableau**

INTERNATIONAL PUBLICATIONS

- "**A Novel Approach for Intelligent Crime Pattern Discovery and Prediction**" Proceedings of the "International Conference on Advanced Communication Control and Computing Technologies-ICACCCT-2016" 978-1-4799-3913-8 - IEEE/ICACCCT2016
- "**Huskysort**"- Improvised sorting algorithm using array access for large scale data types, submitted to (ACDA 2021) - Preprint

TECHNICAL SKILLS AND CERTIFICATIONS

- **Programming/Query Languages:**Python, R, Shiny, Flask, Java, SQL, NOSQL, nodejs, MongoDb, pySpark, C#, Hive, Teradata
- **Modeling and Analysis Tools:** PySpark, PyTorch, RStudio, Hadoop, Apache kafka, AWS (S3, EC2, Sage maker, EMR), Keras, Spacy, DeepNeuralNet, NLTK, Scikit, TensorFlow, OpenCV, Numpy, Pandas, Tableau, Git, Atlassian Tools, Azure, JIRA
- **Certifications:** Machine Learning (Andrew Ng, Stanford), Big Data Analytics (HPES)