DHARINI PARGUNAN

dharinipargunan0995@gmail.com **+1- (972) - 977- 2271** https://www.linkedin.com/in/dharini-pargunan

Education

Masters in Applied Science Engineering - Data Science — Aug 2021 – Dec 2022
The State University of New York at Buffalo, USA — **3.852**

Bachelor's Degree in Information Technology — April 2013 – May 2017
Anna University, Chennai, India — **3.116**

Relevant Coursework

Statistical Data Mining and analysis, Machine Learning (Linear Regression, Logistic Regression, Support Vector Machine, Ridge, Lasso, Naïve Bayes, Unsupervised learning, Decision Trees, Neural Networks (CNN, ANN, RNN)), Clustering Algorithms, Ensemble Methods, Bagged (Random forest, LGBM) and Boosted (LGBM, XGBoost) trees, Multivariate Analysis, Data Structures, Deep Learning, NLP, Predictive modelling, Statistics, Probability, Linear Algebra.
Data Analysis Skills: Data Analytics, Data Manipulation, Data Pre-processing, Data Cleaning, Data Visualizations, ETL, Data Warehouse.

Certifications

Tableau Desktop Specialist, AWS Cloud Practitioner, AWS Machine Learning Specialist, Deep learning Specialization (Coursera**)**

Technical Skills

Programming Languages: Python, R, C, C++, MATLAB, SQL**Data Visualizations :** Power BI, Tableau, MS Excel
Deep learning/NLP Tools: Scapy, NLTK, Tensorflow, Pytorch,Keras, Docker, Kubernetes, Hugging face
Databases: PostgreSQL, SQLite, MySQL, NoSQL, SQL Server, S3
Big Data: Hadoop, Apache Spark, Hive, Databricks
ML Libraries: Scikit-Learn, Pandas, NumPy, Matplotlib, Stats Model, Plotly, Seaborn, Streamlit, Flask, Scipy

Projects https://github.com/dharinipargunan

Customer Churn Prediction* | *ANN* | *TensorFlow* | *Keras* | *Python* |** — ***Apr 2022

- Performed Data cleaning, Data manipulation and pre-processed the Bank customers data. Converted to numerical data using one-hot-encoding and performed normalization. Implemented Artificial Neural Network and built customer churn prediction.
- Evaluated the model using various metrics like precision, recall and accuracy using Confusion Matrix and Classification Report. Achieved an accuracy of ***78.9%*** and a precision score of **86.5%.**

Spam Classification* | *NLP* | *TF-IDF* | *Streamlit* | *Python* | *NumPy* | *Seaborn* |** — ***Mar 2022

- Performed Data-cleaning & Data-Manipulation techniques by removing duplicate rows, handling missing values, and label encoding. Implemented **Feature Generation** and created new features and performed **Exploratory data analysis** on the new features.
- Built **Naïve Bayes Model** (Gaussian, Multinomial, Bernoulli) using Count Vectorizers & TF-IDF and predicted the spam messages.
- Evaluated these models using the metrics **Precision** and **Accuracy**. Achieved an overall accuracy of ***95.93%*** in Multinomial NB
- Created a UI using the **Streamlit** library for the user input & displayed whether it is spam or not.

Movie Recommendation System using* | *ALS* | *Pyspark* |** — ***Feb 2022

- Built a movie recommendation system using ALS in Spark programming and Movielens 100K dataset. Ratings of the movie are predicted for the movies that were not predicted. Implemented K- Means clustering to recommend the movies based on the ratings.

House Price Prediction |Python | Flask | Linear Regression | Lasso | Ridge| — ***Jan 2022***

- Performed data cleaning techniques on real estate house price data & pre-processed the data & Implemented Linear regression models.
- Evaluated the model using MAE, R-squared error, RMSE and achieved an accuracy of **97.78%** on the test data

Credit Card Customer Segmentation* |R Studio | R Shiny|** — ***Dec 2021

- Classified customers based on behavioral variables from credit card transactional data of customers which helps to identify the potential customers to purchase credit cards and helps various industries to devise a marketing strategy.
- Pre-processed the data and found the Optimal value of K using the **Elbow method**, **Gap statistics** for clustering the customers and implemented K-means and Hierarchical Clustering models. Used R shiny dashboard to give the detailed view and the flexibility of choosing the clusters for exploratory data analysis.

Experience

Data Scientist Associate – State Bank of India, *Chennai, India* — ***Dec 2018 – May 2021***

- Applied data mining techniques, performed statistical analysis, and built high quality loan prediction machine learning models which increased the loan availing percentage of the customers by **10% in the next quarter.**
- Ensured data quality and data validation throughout all stages of acquisition and processing, including areas such as data collection, ground truth generation, normalization, and transformation.
- Analyzed customer data and implemented Clustering models such as K-means, Hierarchical models to find different groups of customers and identified the risk level, which further increased the sales of cross-selling financial products by **33%.**
- Created a machine learning model to predict customer churn based on historical data, reducing the company's churn rate by **5**%.
- Built and maintained distributed machine learning pipelines. Built ETL pipelines for product and marketplace metrics.
- Improved the accuracy of email classification from **85% to 95%** through feature engineering and regularization techniques
- Forecasted the increase income trends for future months using time series analysis such as ARIMA, SARIMA.
- Collaborated with other team members on data analysis workflow, data pipelines, analysis techniques and for model improvements.

Data Analyst – State Bank of India, *Chennai, India* — ***Aug 2017 – Dec 2018***

- Worked independently and as a team for internal business problems leveraging large amount of internal and external data to generate insights and developed business recommendations which increased the net profit of banking products by **20%** in the **year 2020.**
- Worked on Finance dataset and converted data into actionable insights by predicting and modelling the future outcomes on data history.
- Implemented pricing experiment to perform cohort analysis that identified an opportunity to improve the customer lifetime value by **21%**
- Prepared various BI dashboards in **Power BI** for month-on-month credit/insurance business reports for different **KPI** measurements
- Implemented statistical analysis, data wrangling using Python, and created visualizations/ dashboards and stories using **Tableau**
- Evaluated and analyzed the insurance/loan claim datasets using **Excel V-Lookup,** Chi-Square, Normal and T-Distribution
- Extracted and cleaned 3M rows of Banking customers data from **SQL Server DB** and created intuitive dashboard using **Power BI.**
- Created **ETL pipelines** using **Tableau Prep Builder** for data manipulation & for loading the dataset from Excel, CSV files, and SQL database into tableau.