DHARINI PARGUNAN

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

Education

Master's 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, Random Forest, Neural Networks (CNN, ANN, RNN)), Clustering (K-means, Hierarchical), Algorithms, Ensemble Methods, Multivariate Analysis, Data Structures, Deep Learning, Natural Language Processing, Predictive modelling, Mathematics, 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	**Databases:** Oracle 10G, SQLite, MySQL
Data Visualizations : Power BI, Tableau, MS Excel	**Big Data:** Hadoop, Apache Spark, Hive, Databricks
Deep learning/NLP Tools: Scapy, NLTK, Tensorflow, Pytorch, Keras, Docker, Kubernetes	**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, Multinominal, 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| *Jan 2022*
- 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 the Optimal value of K is found using the **Elbow method**, **Gap statistics** for clustering the customers using K-means Clustering and Hierarchical Clustering. Used R shiny dashboard which gives the company the flexibility of choosing the clusters and exploratory data analysis.

Experience

Data Scientist Associate – State Bank of India   **Dec 2018 – May 2021**
- Applied data mining techniques, performed statistical analysis, and built high quality loan prediction systems 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.
- Managed EDA (Exploratory Data Analysis), Data Cleansing, Data Manipulation, Data per-processing and Data munging to feed the data into a machine learning model and hyper-tuned the parameters to achieve a better accuracy.
- 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.
- Experienced in Python Scripting and PostgreSQL and built APIs. Collaborated with other team members on data analysis workflow, data pipelines, analysis techniques and for model improvements.

Data Analyst – State Bank of India   *Aug 2017 – Dec 2018*
- Analyzed financial data and worked on the development of new financial models and presented detailed financial forecasts and budgets.
- Monitored expenses, and revenue; conducted weekly reports and presented findings and results to managers and executives using the **Power BI** tool. Built intuitive interfaces, infographics, and visualizations to tell stories with data in **Tableau.**
- Worked across marketing and business units to build out the reporting infrastructure in Tableau.
- Build data visualizations using **SQL** and **Tableau** for business KPIs that reduced manual reporting by 9 hours weekly.
- Prepared various BI dashboards in Tableau for month-on-month credit/insurance business reports
- Evaluation and analysis of insurance/loan claim datasets using **Excel V-Lookup, Chi-Square, Normal and T-Distribution**