Kankshi Dhar

University of Maryland, College Park **May 2020**
kankshi01@outlook.com github.com/kankshidhar 857-241-7195

Education

University of Maryland - Master of Information Management (STEM) **College Park, MD**
Data Analytics, Information Retrieval Systems, Data Science Models, Database Design *Aug 2018-May 2020*

University of Mumbai - Bachelor of Engineering in Computer Science **Mumbai, MH**
Relational Databases, Data Structures, Neural Networks, Artificial Intelligence *Aug 2012-June 2016*

Key Skills

- Programming Languages: Python, R, MySQL, Java, C++, Shell Scripting, NoSQL, HiveQL,PostgreSQL
- Software: Nltk, Tensorflow, Keras, Sklearn, Pandas, tidyverse, Statsmodels, Caret, dplyr, Hadoop, MapReduce
- Tools: Tableau, Power BI, H20, ArcGIS, Git, Jira, Confluence, AWS Redshift, Sharepoint, MATLAB, SAS EG
- Machine Learning: Random Forests, Logistic regression, SVMs, Naive Baiyes, Neural Networks, Clustering, Tensorflow
- MS Office Suite: MS Word, MS Excel, MS Access, MS Powerpoint

Work Experience

CrowdDoing.World , Data Scientist *July 2020 - Present*
- Building a Hybrid recommendation system for Medicinal foods based on classifications of disease, type of disorder & consumer ratings
- Implemented classification with topic modeling of customer reviews & regression model for aggregate table value recommendations

University of Maryland, Data Analyst *Aug 2018 - May 2020*
- Reduced run-time by **26%** & enhanced data processing by creating complex SQL queries to generate analytical insights for stakeholders
- Built ad-hoc reports with multifaceted data analysis for business teams leveraging BI tools such as SAS EG & Oracle Hyperion BI
- Optimized hyper parameters with historical data for University metrics, boosting trend analysis using Python3 - seaborn & Tableau
- Spearheaded FY19 and FY20 University financial impact profile creation with quantitative analysis using over **10 million records each**
- Work with stakeholders throughout the organization to identify opportunities for leveraging company data to drive business solutions

Cheniere Energy Inc., Information Analyst Intern *May 2019 - Aug 2019*
- Reduced run-time by **6 hours** by creating a Unique data source & Automating weekly volume report generation using NumPy & Pandas
- Established KPIs to monitor the compliance of processes and associated records to company standard operating procedure requirements
- Performed data mining to develop a tool with data validation, tracking & forecasting of Budget trends by quantitative analysis using multivariate regression & Support Vector Machines with Python libraries - scikit learn, Pandas, plotly

Reliance Jio Infocomm Ltd., SAP CRM Developer *Sep 2016 - Jun 2018*
- Lead to the organization **ranking #1** across the country by developing Enterprise IOIP services in a cross functional team to launch Individual owned products using ERP Suites such as SQL & NoSQL queries for over **30 million customers**
- Designed & Managed databases with a focus on performance, security, scalability and disaster recovery for large Enterprise Customers
- Analyzed customer preferences data & developed 'Lead' project to increase profitability by reducing churn rate by **11%**
- Created Data lakes & complex SQL queries for AWS RedShift & S3 bucket on partner codes decreasing run-time by **4 hours**
- Performed unit & integration testing of application for gap assessments to assess cost/benefit and technical viability of RDS solutions

Research Projects

Text Mining & Classification for HireSmith Career Portal *Jan 2020 – May 2020*
- Performing text mining & classification on Job descriptions for skill trend analysis in multiple industries within past 10 years
- Pre-processed data & applied Term Frequency-Inverse Document frequency with Topic modeling to analyze job description and functions
- Identified 16 Industries & examined it's independent effect with 7 other categorised attributes across 5 million records
- Implemented Animated Tableau dashboards and plotly to display categorised attribute's trends across years

Statistical Analysis on Ocean Waste Decomposition *Jan 2020*
- Created a time series analysis of Ocean waste collected & it's decomposition rates to project sustainable development insights
- Pre-processed data to eliminate inconsistencies, normalize numerical variables & treat contextual anomalies using AWS ESRI ArcGIS
- Analyzed the categories of trash, decomposition rate & correlation to cleanup participants using statsmodels package

Medicare Rate Forecasting Model *Aug 2019 - Dec 2019*
- Achieved 12% accuracy over previous models by developing logistic regression & Neural network model with ensemble methods to detect correlation, patterns & predict average Medicare provider's reimbursement rate in a dataset of 10 million records
- Conducted data pre-processing & augmentation to eliminate outliers, avoid overfitting & data exploration using Plotly, sklearn
- Models verified that highest paid provider will be plastic surgeons, Cardio-thoracic surgeons and Anesthesiologists in FY20

Awards & Achievements

- Publication : "Offline Signature Corroboration Using Gray Level Features". International Journal of Engineering Science. Issue 3421
- Awarded Best Global Sustainability Insight Award at the "Data Challenge 2020" for Ocean Cleanup Data set by Booz Allen Hamilton