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   **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
○ Software: Nltk, Tensorflow, Keras, Sklearn, Pandas, tidyverse, Statsmodels, Caret, dplyr
○ Big Data: Hadoop, MapReduce, Spark, Hive, PostgreSQL
○ Tools: Tableau, Power BI, H20, ArcGIS, Git, Jira, Confluence, AWS Redshift, Sharepoint, MATLAB
○ Analytics: SAS Enterprise Guide, Oracle Hyperion BI, Oracle DB, Clear Analytics, SAP BI, Visio, Stata

Work Experience

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 & multifaceted requirements 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 Financial impact profile creation for University trend 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*

○ Boosted run-time by **6 hours** by creating a Unique data source & Automating weekly volume report generation using NumPy & Pandas
○ Created & Maintained Data governance risk assessment records w.r.t PHMSA for FY20 using Advanced Excel, Visio & Pivot Charts
○ Developed a reusable report to track & forecast Budget trends using NumPy, matplotlib & Pandas for Logistic regression models
○ Established KPIs to monitor the compliance of processes and associated records to company standard operating procedure requirements

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 SAP CRM, SAP HANA 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 & developed 'Lead' project to increase profitability by reducing churn rate by **11%**
○ Increased performance by **18%** on AWS RedShift & S3 by performance tuning and debugging existing ETL processes on partner codes
○ 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 TF-IDF with Topic modelling to analyze full text of 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 with Best Sustainability Insight Award at the "Data Challenge 2020" for Ocean Cleanup Dataset by Booz Allen Hamilton
○ Elected Director of Public Relations at Student Committee, 'YUVA' at University of Maryland (FY2019)