DINESH REDDY VANGUMALLI

dvangu01@villanova.edu • ☎ 4849194772
LinkedIn

EDUCATION

Villanova University, USA	MS in Applied Statistics	May 2019
The University of Manchester, UK	MS in Business Analytics	Dec 2015
Acharya Nagarjuna University, India	Bachelor of Technology, Electronics & Communication Engineering	Apr 2008

PROFESSIONAL EXPERIENCE

GlaxoSmithKline (via Atrium), Philadelphia, USA – *Data Scientist, AI and Text Analytics Co-Op* **Sep 2019 – Present**

- Analyzed unstructured data from multiple sources including **Voice of Customer** (VoC) feedback using **Natural Language Processing** and delivered impactful visualizations using **Power BI** to identify opportunities in product development
- Implemented **NLP** models such as **text classification, named entity recognition** using **spaCy** and **topic modeling** using **quanteda** and **genism** packages to construct compelling narratives for better consumer experience

Oracle America, Inc., Conshohocken, USA – *Data Science Co-Op* **Jan 2019 – Aug 2019**

- Analyzed Construction and Engineering data using **SQL, R** and **Python**, to **uncover trends** and **derive actionable insights** in the project operations and helped drive efficiency at jobsites
- Developed an **Anomaly Detection model** that identifies outliers in the activities of a project using **Machine Learning algorithms** such as **Isolation Forest** to improve strategic and project-level decision making
- Identified and **re-engineered features** in data, which led to **35% improvement** in the already existing ML models

Dealmaar Technologies, Hyderabad, INDIA – *Data Analyst, Business Intelligence* **Feb 2016 – Jul 2016**

- Extracted, cleaned, and manipulated large-scale datasets using **SQL**, **R** and **Python** to prepare for exploratory data analysis and built **BI dashboards** using **Tableau** to present insights to stakeholders
- Designed **test strategies** for sellers enabling them to make **better investment decisions** and optimize spend by evaluating category performance, generating an **incremental lift** of **26%** in sales
- Guided enterprise wide marketing strategies by applying **PCA** and **k-Means clustering** techniques to high-dimensional data and come up with **customer segmentation;** delivered **$800K incremental sales lift**

LAD Software Solutions, Hyderabad, INDIA – *Data Analyst, Customer Insights* **Jul 2010 – Dec 2013**

- Analyzed sales data of a travel company using **pandas, scikit-learn** and **built a predictive model** to identify repeat customers, enabling marketing team to recommend a business strategy to improve the repeat rate
- Investigated transaction level datasets along with demographic data, using **SQL** and **R,** to provide **customer insights** and generated **ad-hoc reports** for optimizing business performance
- Built statistically valid tests (**A/B, Hypothesis and ANOVA tests**) to evaluate impact for landing page analysis and marketing campaign effectiveness on key metrics such as visits, spend, etc.

PROJECTS

Predicting Bone Age from Radiographs using Deep Learning

- Developed a **Convolutional Neural Network (CNN)** algorithm using **AWS**, **Google Cloud** instances and frameworks **CUDA, TensorFlow** and **Keras**, to determine skeletal age from pediatric hand radiographs; achieved **Mean Absolute Error** (MAE) of **7.2 months**, better than the predictions of doctors (8 months)

ETL Pipeline for a Data Lake using Apache Spark and AWS

- Built and scaled an **ETL pipeline** that extracts data in JSON format from **S3 buckets**, processed using **Spark** and load data back to S3 as **dimensional tables** using efficient partitioning and parquet formatting

TECHNICAL SKILLS

Languages: Python, R, SQL, SAS
Databases: Oracle, MS SQL Server
Visualization: Tableau, QlikSense, seaborn, ggplot2, OBIEE, Power BI
NLP: gensim, spaCy, quanteda

Machine Learning: Regression, Decision Trees, PCA, Random Forest, XGBoost, k-Means, k-NN
Deep Learning: TensorFlow, Keras, PyTorch, ANN, CNN, Autoencoders

Big Data: Spark SQL, PySpark, Spark MLlib, SparkML, Hive
Cloud: AWS, S3, EC2, Redshift, EMR, Google Cloud, BigQuery
Additional Tools: Git, Linux, Jira