PRANEETH BOMMA

+1 (980)-318-6668 | praneeth.bomma401@gmail.com | https://www.linkedin.com/in/pbomma | https://github.com/pbomma | http://praneethbomma.com/

Graduate student seeking a full-time position with 2 years of Academic and Professional experience in Data Science & Analytics

---

EDUCATION

Master of Science in Information Technology(Advanced Data and Knowledge Discovery), GPA 4/4   **Jan 2017-May 2018**
The University of North Carolina at Charlotte, North Carolina

Bachelor of Technology in Electronics and Communications,GPA 3.5/4   **Jul 2012 - Aug 2016**
Jawaharlal Nehru Technological University – Hyderabad, India

TECHNICAL SKILLS

Programming:	SQL, R, Python, Core Java, SAS, Scala, Pig, Hive
BI/Analytics Tools:	MS Excel, Tableau, WEKA, Google Analytics, Looker
Statistics/Machine Learning:	Regression, ANOVA, Clustering, PCA, NLP, Sentimental Analysis, LDA, Bayesian Statistics, Random Forest, Time-series analysis
Web Languages:	HTML5, CSS, jQuery, JavaScript
Databases:	MySQL, Oracle SQL, MS SQL, Postgresql
IDE:	Eclipse, Microsoft Visual Studio
Cloud Services:	Amazon Web Services, Microsoft Azure, Google Cloud, Cloudera

---

PROFESSIONAL EXPERIENCE

DATA ANALYST INTERN, *Informative Technologies Inc, Charlotte, NC, USA*   **Jan 2018 - Present**

- Created Entity relationship (ER) diagrams and developed database objects such as tables, views, indexes, triggers & stored procedures for a fashion industry based start-up inspired by design concepts from Pinterest, Instagram, and Tinder.
- Applied query optimisation techniques for tuning the query by creating indexes to improve the Query & System performance.
- Queried data for analysis from high relational databases by writing complex SQL queries. Processed and verified the integrity of data used for analysis, came up with ad-hoc reports on daily basis and presented the data in charts, graphs and tables.
- Identified data quality issues and used MS EXCEL to audit data using pivot tables, vlookup, formulas, conditional formatting across all the file types. Performed the batch processing of data, designed the SQL scripts, control files, batch file for data loading.
- Involved in various phases of Software Development Life Cycle (SDLC) and Agile/Scrum methodologies including Analysis, Design, Testing, Implementation and Maintenance. Experienced in AWS services like EC2, RDS and S3.

ACADEMIC PROJECTS

Finding Surprising Documents on Online Health Information| Tools Used: *R*

- Developed a computational approach using R by applying association mining, text modelling, and machine learning techniques to identify "surprising" news from a news corpus related to diabetes.
- Performed clustering analysis (K-Means, K-Medoid, SK-Means), PAM, cosine similarity, and word cloud operations on text corpus.

Hire Heroes USA- Client Management| Tools Used: *SAS, R, Excel and Tableau*

- Developed a Multiple Regression Model which explains 27% of the variation in how quickly a veteran gets hired.
- Used Text mining to categorize data into text topics and decision trees to explain correlation between client & employer's demographic data.

Spatial and Time-Series Analysis of SFO Crimes| Tools Used: Python, Matplotlib, Pandas, Seaborn, Numpy and ARIMA model

- Performed spatial distribution over time and time series analysis for a 15 year dataset of reported incidents from SFPD.
- Trained and fine-tuned an ARIMA model to forecast the number of theft incidents per month.

Lending Club - Loan Status Prediction (Kaggle)| Tools Used: Python, Scikit Learn, Pandas

- Performed feature selection, extraction, built classification and ensemble methods to predict borrowers who tend to default.
- Applied k-fold cross validation to select best parameters of model and obtained 91% prediction accuracy using ensemble methods.

An Electronic Medical Record for an Outpatient Clinic| Tools Used: *MySQL*

- Designed and developed a complete OLTP database for an Outpatient Clinic that can efficiently store, retrieve, manipulate, and query records. Wrote Triggers and Stored Procedures to capture updated and deleted data from OLTP systems.
- Implemented Authentication and Role based access control to all the data tables & used views and indexes for easy data access

Twitter Text Analysis – Movie Success| Tools Used: *Python, Tweepy API, NLTK*

- Tweets crawled using the Tweepy API in Python were pre-processed to create a corpus for analysis using NLTK module.
- Performed sentiment analysis & created a tag cloud of top 50 words in tweets to understand the audience sentiments about movie.

CERTIFICATIONS

- Coursera Certifications: SQL for Data Science, Getting and Cleaning Data, R Programming, Mastering Data Analysis in Excel
- Google Analytics: Google Analytics Individual Qualification, Advanced Google Analytics
- Oracle Certified Java SE6 Programmer