KRISHNA GOLLAPUDI

https://www.linkedin.com/in/gkrishna19/ | http://www.krishnagollapudi.com/ | https://github.com/KrishnaGolllapudi
email: vgollap1@uncc.edu | phone: (+1) 980-598-0789

Summary:

Data Science Graduate with professional and research experience in Data Science and Data Analytics field with expertise in Machine Learning, Data Mining, Statistical Modeling, Time Series, Information Visualization, Text Analytics, Natural Language Processing. Proficient in Python, R, Tableau, SQL with the ability to translate given data providing meaningful insights solving business problems.

EDUCATION:

Master of Science: **Information Technology, GPA: 3.9** **Jan 2017 - May 2018**
University of North Carolina at Charlotte – Charlotte, North Carolina

Bachelor of Engineering: **Electronics and Communications, GPA: 3.5** **Aug 2011 - May 2015**
Osmania University – Hyderabad, India

WORK EXPERIENCE:

PROLYTICS LLC, Charlotte, North Carolina, USA. **– Data Scientist Intern** **Jan 2018 – Apr 2018**

- Performed advanced statistical analysis and predictive modeling on MLB, NBA Draft data (3 years of NBA data) and NCAA college stats
- Adapted **Machine Learning** algorithms in predicting player's match up analysis based on their game position, historical NBA data
- **Data wrangling** of MLB data and **predictive analysis** on the performance of each player in future matches by **classification models**
- Developed an LSTM RNN to project NBA player's expected performance in the draft over 2 years and visualized the prediction curve
- Performed **sentimental analysis** on player's twitter data to determine whether he had a positive or negative event prior to the game

ACCENTURE, Hyderabad, India. **– Associate Software Engineer** **Mar 2016 - Dec 2016**

- Contributed as an active team member with Global Resource Management project for client: **Microsoft** with **Agile** methodology
- Modified the web test scripts according to the API changes and created pipelines in Azure Data Factory(**ADF**), SQL Jobs
- Executed constant/load tests in **Microsoft Azure** which include performance monitoring, performance test analysis, performance tuning
- Assisted in **visualization** of the employee's demographic data and understanding the budget problems in resource management
- Presented detailed weekly reports about the performance of the dashboard and help the team identify areas of improvement

HCL CDC, Hyderabad, India**. – Software Engineer Intern** **Dec 2015 – Feb 2016**

- Worked on an Internal project of HCL for Customer Query Tracking System
- Responsibilities include system design, creating E-R diagram, tables and stored procedures

INFOSYS LIMITED, Mysore, India. **– Systems Engineer Trainee** **Jun 2015 – Nov 2015**

- Trained on PYTHON, JAVA, HTML, CSS3, JavaScript, Oracle SQL
- Designed and developed an SQL Database system for an internal Business Enterprise Application

TECHNICAL SKILLS:

Programming Languages/Tools:	Python, R, Tableau, SQL, SAS, Java, C++, MS Excel, Google Analytics, QlikView, Weka, GitHub
Machine Learning/Statistics:	Regression, Classification, Rule Mining, ANOVA, NLP, Text Analysis, Bayesian statistics, TF-IDF, SVM, PCA, LDA Information Retrieval, Forecasting, Survival Analysis, Time-Series Analysis, RNN, LSTM, Deep Learning
Libraries (Python and R):	pandas, scikit-learn, scipy, numpy, matplotlib, xgboost, cluster, tidyverse, nltk, word2vec, genism, shiny, statsmodels, seaborn, plotly, bokeh, arima, tensor flow, convolution2D, H2O**,** pytorch, Theano, keras
Database/Big Data:	MySQL, MongoDB, Oracle SQL, SQL/PL, Teradata, MS SQL Server, Hadoop, Spark, Scala, Pig, Hive
Cloud Services:	Amazon Web Services, Microsoft Azure, Google Cloud, Digital Ocean, Horton works, Cloudera

PROJECTS:

Real Estate Housing Prices Predictions| Techniques: Feature Scaling, k-fold, Gradient Boost, Grid search**| Tools:** ***Python, Tableau***

- Performed exploratory data analysis, feature scaling, k-fold cross validation and grid search to achieve the most approximate prediction
- Achieved the least mean square error in predicting housing prices of King county housing data using Gradient Boosting

Loan Default Prediction| Techniques: feature selection, feature extraction, classification**| Tools:** ***Python***

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

Surprising Discoveries on Diabetes Data| Techniques: Tokenizing**,** Clustering, Cosine Similarity, PAM & Word Cloud, SK-means**| Tools:** ***R***

- Developed a computational approach using R programming to identify "surprising" pattern from **diabetes** related news corpus
- Applied the unsupervised machine learning techniques to achieve the surprising discovery from given text corpus of 10000 documents

Twitter Text Analysis – Movie Success| Techniques: NLTK, sentimental analysis, tag cloud**| Tools:** ***Python, NLTK, matplotlib, TweepyAPI***

- Tweets crawled using the TweepyAPI in Python were pre-processed to create a corpus for analysis using NLTK module
- Performed sentimental analysis and created a tag cloud of top 50 words in the tweets to estimate audience sentiments about the movie

Geo Spatial Analysis and Time-Series Forecasting of SFO Crimes| Techniques: EDA**,** Time Series Analysis, Spatial Analysis**| Tools:** ***Python***

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