

AAKARSH NADELLA

875 Lockefield St, Indianapolis, IN 46202 | (469) 900-9835 | aakarshnadella14@gmail.com |
www.linkedin.com/in/aakarshnadella | <https://github.com/aakarshnadella>

EDUCATION

- **Master of Science in Computer and Information Science** GPA: 3.6/4
Indiana University-Purdue University Indianapolis, IN May 2019
- **Bachelor of Technology in Computer Science and Engineering** GPA: 8.89/10
Jawaharlal Nehru Technological University Kakinada, India May 2017

TECHNICAL SKILLS

- **Programming Languages** : Python, Java, R programming, C++, C, HTML, CSS, JavaScript, D3.js
- **Query Language** : MS SQL, PL/SQL, Apache Pig, PySpark, Apache Spark, MongoDB
- **Machine Learning Libraries** : NumPy, Pandas, Scikit-Learn, TensorFlow, NLTK, Keras, Matplotlib
- **Tools and Technologies** : ERwin, SQL Server Management Studio, Tableau, Git/GitHub, DBeaver, Jupyter

PROFESSIONAL EXPERIENCE

Indiana University | Indianapolis, Indiana Sep'2017 – May 2019
Research Assistant

Multi-class Classification of Biomedical data to find Protein/Gene Interaction

- Adapted and deployed several Natural Language Processing concepts such as POS tagging, Named Entity Recognition, Regular Expressions etc. to determine context words in sentence
- Hosted UMLS database containing genes and proteins on local SQL server and established its connection with Python to eliminate those sentences which do not contain protein or gene names, thereby improving data quality
- Produced word embeddings on processed dataset using one-hot encoding and Word2Vec
- Built 2 deep learning models, one with Convolutional Neural Networks (CNN) and another with combination of both Recurrent Neural Networks (RNN) and CNN to classify given data into 10 different classes using Python
- Utilized several modules such as TensorFlow, keras, Scikit-learn, Matplotlib, NumPy, Pandas, NLTK, etc. in Python

Sentiment Analysis of Indy 500 Twitter data

- Collected data using Python by connecting to Twitter Streaming API & built model using Long Short-Term Memory (LSTM) architecture based RNN
- Built Naïve Bayes Classifier using Bigrams after cleaning reviews, eliminating stop words and performing stemming
- Compared performance of both models, generating Confusion Matrix, Receiver Operating Characteristic curve (ROC) and Normalized Root Mean Square Error (NRMSE)
- Summarized inconvenience experienced by audience based on negative reviews using Abstraction-based summarization
- Utilized TensorFlow, NLTK, Genism, Matplotlib and ggplot packages in Python

M3Bi India Pvt. Ltd. | Hyderabad, India May 2016 – July 2016
Database Intern

- Conducted detailed functional dependency analysis and designed database according to client's requirements, achieving presumably highest level of Normalization (Boyce-Codd Normal Form)
- Designed database in ERwin, generated prototype using SQL Server 2016 and efficiently loaded records in tables
- Ensured security by restricting user's access, constraining data using triggers, stored procedures and views using MySQL and sensitive data is protected using Dynamic Data Masking

PROJECTS

Understanding and Implementing Big Data Concepts using Stack Overflow Data Aug'2018 – Dec'2018

- Managed and analyzed a large data (around 75GB) on Stack overflow site using Spark RDD's transformation, actions, Spark SQL and PIG scripts to observe trends with users and questions posted. Wrote an XML parser in Python
- Utilized Amazon EMR, Amazon EC2 and Amazon S3 for processing data, then visualized results using Tableau

Analyzing Performance of Machine Learning Algorithms using Walmart Stock Price Data Jan'2018 – May'2018

- Performed univariate, multivariate feature selection to figure out optimal feature set and cross-validated it with Recursive Feature Elimination (RFE)
- Built predictive models using machine learning algorithms such as Auto Regressive Integrated Moving Average (ARIMA), Support Vector Machines (SVM) and RNN using R programming
- Computed Confusion Matrix and NRMSE to measure performance and remodeled models to improve accuracy

Predictive Analytics on Tax Evasion Using Feature Engineering Strategies Aug'2017 – Dec'2017

- Performed feature selection to figure out optimal feature set for data analysis and built predictive models using Random Forest, Support Vector Machines (SVM), linear regression & Back Propagation Neural Networks using R programming
- Evaluated model's performance using NRMSE and Coefficient of Determination (COD)