Amit Kumar Yadav

amitsyadav.yadav716@gmail.com | (812) 351-9502 | www.linkedin.com/in/amit--yadav | www.github.com/savewater1/

EDUCATION

Indiana University, Bloomington, Indiana — *August 2018 – May 2020*
Master of Science in Data Science — GPA: 3.84/4.0

Gujarat Technological University, India — *August 2013 – June 2017*
Bachelor of Science in Computer Science — GPA: 3.7/4.0
Relevant Coursework: Calculus and Linear Algebra, Database Management System, Data Structures, Analysis and Design of Algorithms, Data Compression and Retrieval, Data Mining and Business Intelligence, Artificial Intelligence, Python

TECHNICAL SKILLS

Programming Languages: Python *(SciPy, NumPy, pandas, scikit-learn, matplotlib, plotly)*, SQL, R, C++, Java, C, Scala
Databases: Relational Databases (PostgreSQL, MySQL, MSSQL Server), Amazon RedShift, MongoDB
Operating System: Linux (Ubuntu, CentOS, Red Hat), Windows, Mac OS
Deep Learning: TensorFlow, Keras, PyTorch
Tools: GIT, Heap Analytics, Mode Analytics, Tableau, PySpark, Spark, Jupyter, Docker, Kubernetes
Additional Skills: Big Data systems, Hadoop, Parallel computing, AWS, Google Cloud, ETL, Data pipelines, Distributed Systems, Statistical Hypothesis Testing, Statistical Analysis, Data Cleaning, Data Modeling, Forecasting, Qualitative Analysis, Quantitative Analysis, Data Visualization

WORK EXPERIENCE

Supply Clinic Inc., Chicago, IL | Python, PostgreSQL, Redshift, Heap Analytics, Selenium, Beautiful Soup — *June 2019 – August 2019*
Data Intern

- Engineered web-scrappers for data collection and ETL pipelines for data integration to create a database of products - used for improving marketing campaigns, operations, customer acquisition, customer engagement and cost negotiations
- Streamlined data analysis process to efficiently query databases and enable collaborative analysis of user site behavior and purchase patterns; utilized to revamp UI/UX design, conduct A/B testing, and improve user experience
- Interpret and analyze structured and unstructured data sets for ad-hoc analysis and to generate actionable insights
- Collaborated with data warehouse team and utilized data repositories for trend (time series) analysis and modeling of customer needs and behavior, resulting in improved demand forecasting and superior user engagement
- Created and automated delivery of visually impactful dashboards to report findings and monitor performance metrics

PROJECTS

Predictive Analysis on Energy Consumption | Exploratory Analysis, LightGBM, statsmodel — *November 2019 – January 2019*

- Conducted thorough explanatory data analysis, engineered new features, and built linear models and robust LightGBM model to facilitate predictive analytics on the energy consumption of buildings given datasets containing meter readings, building metadata and local weather data

Distributed Big-Data Processing System | python, GCP, RPC, Map-Reduce — *September 2019 - November 2019*

- Engineered a persistent key-value store, inspired from Memcached, running on remote server, providing concurrency and ACID transaction facilities
- Developed a scalable and fault-tolerant *MapReduce library* which exposes API using remote procedure calls (RPC) to provision gcloud compute instances and submit Map-Reduce tasks. The system, like Hadoop, supports local, standalone, and fully distributed modes and allows user to create custom Map and Reduce tasks

Distress Detection from Speech Signal | python, librosa, keras, CNN, LSTM, ensemble — *March 2019 – April 2019*

- Researched spectral and temporal speech features and used them in an ensemble of deep learning models comprised of CNN and/or LSTM layers to perform classification of emotional state of an individual from speech signal
- Created a system that can detect distress in speech with accuracy of 75% and further distinguish the distressed, classified by emotion of anger, from the distressed, classified by emotion of fear, with an accuracy of 90%

Sentiment Analysis on Amazon Review Data | python, scikit-learn, TF-IDF, ensemble — *November 2018 – December 2018*

- Employed TFIDF feature vectors to design a system that classifies reviews into one of 5 classes (ratings) using ensemble of 9 weak classifiers (Logistic Regression, SVM, etc.) which achieved 35% improvement in accuracy from baseline
- Experimented with various sampling techniques to balance the dataset and researched techniques to incorporate lexicographical features

Mini Projects — *August 2018 - Present*

- *Analysis of tweets* from FIFA world cup 2018 using MongoDB and spark's DataFrame API
- *Spam Detection* – Classified emails as spam or ham using Bag of Words model and Naïve Bayes classifier
- *Keyword Extraction from Review Data* using ELMo Word Embeddings and clustering techniques
- Developed *Variational AutoEncoder* for MNIST dataset to obtain latent representation, used to construct new images of digits

TravelBug | JAVA, Hibernate, REST, MS-SQL — *August 2016 – February 2017*

- Designed a JAVA based web application using REST architecture, Spring DAO pattern and Hibernate ORM, aimed at providing a wholesome system for registering, searching, booking, canceling, etc. of the tours and travel packages where users have the flexibility to customize their own packages or select from pre-existing packages