SAMUJJWAAL DEY

+1-224-484-9321 | deysamujjwal@gmail.com | LinkedIn | GitHub

EDUCATION

Master of Science in **Computer Science |** *University of Illinois at Chicago (UIC), Illinois* Aug 2019 – May 2021
Coursework: Introduction to Data Science, Information Retrieval, Statistical Natural Language Processing, Deep Learning for Computer Vision, Cloud Computing, Object-Oriented Programming, Knowledge Graphs, Visual Data Science, Computer Algorithms

Bachelor of Engineering in **Computer Engineering |** *University of Mumbai (VESIT), India* Jul 2015 – May 2019
Relevant Coursework: Data Structures, Database Management Systems, Artificial Intelligence, Soft Computing, Data Warehouse and Mining, Software Engineering, Parallel and Distributed Systems

TECHNICAL SKILLS

Languages, databases, software, OS: Python, Scala, Java, C++, R | SQL, MySQL | Docker, Git, Jupyter, Octave | Linux, Windows

Data science: Numpy, Pandas, SciPy | Data visualization (Matplotlib, Seaborn, D3.js, Three.js) | Statistical Modeling | Regression, Classification, Clustering | Hypothesis Testing | Exploratory Data Analysis | Computer Vision (OpenCV)

Machine learning: Scikit-learn | Deep Learning (PyTorch, Keras) | NLP (HuggingFace, Transformers, NLTK, SpaCy)

Others: Cloud (Azure ML, AWS EC2, S3, EMR) | Apache Hadoop, Spark | Akka | Javascript, HTML, CSS | Gradle, sbt | JUnit

SELECTED PROJECTS

Multilingual Chatbot | *Python, Keras, Transformers, TkInter*
- Implemented a conversational multilingual chatbot capable of responding to user queries in more than one language
- Experimented with **Transformer** models mBART, T5 & OPUS-MT for **language detection** and **translation**. Trained a **Keras** Sequential 3-layer **neural network** model using **Stochastic Gradient Descent** optimization

Overlay Network Simulator using Akka | *Scala, Akka, Akka-HTTP, sbt, Docker, ScalaTest*
- Simulated distributed hash tables using **Chord** and **CAN** overlay network algorithms with **Akka actor** as an abstraction for 25 nodes. Incorporated **Akka-HTTP** to expose hash table functions as REST API for **asynchronous read/write** requests
- Containerized the application and runtime dependencies using Docker and deployed on **DockerHub** & **AWS EC2**. Integrated Bitbucket Pipelines/GitHub Actions to automate **CI/CD workflows** for build & deployment

MapReduce on DBLP data | *Scala, Hadoop, sbt, AWS EMR, ScalaTest*
- Leveraged **Apache Hadoop** and **Scala** to parse & analyze **2 million** DBLP publication records using the **MapReduce** framework, and deployed on an **AWS Elastic Map Reduce** cluster
- Performed analytics to identify top authors & publications, and authors & publications with most co-authors at each venue

Cloud Sim Plus Cloud Simulators | *Java, Scala, sbt, ScalaTest*
- Simulated execution of 50 cloudlets on **cloud infrastructure** using Cloud Sim Plus framework. Conceptualized 8 datacenters on a mix of **SaaS**, **PaaS** & **IaaS** architecture models using different policies and constraints for VM allocation and execution
- Evaluated 5 optimal **pricing models** and **load balancing heuristics** to maximize performance at reduced expenses

Web Search Engine on UIC Domain | *Python, Nltk, beautifulsoup*
- Devised a scalable **web crawler** to traverse and retrieve **6,000 web pages** on the UIC domain using a **breadth-first** strategy
- Executed **tokenization** and **stemming** to index **168,833 unique tokens** into a **TF-IDF** vector-space model. Achieved average precision of **90%** for the top 10 most relevant web pages retrieved for search queries

PageRank on WWW conference corpus | *Python, Nltk, NetworkX*
- **Parsed** & loaded each document from the **1,300+** WWW conference abstracts into **undirected word graphs**
- Executed **PageRank** on each word graph & **scored n-grams** formed from adjacent words. Calculated **Mean Reciprocal Rank** for top-k ranked n-grams using an author annotated gold standard

Spam E-mail Classifier | *Python, Scikit-learn, Matplotlib, Pandas*
- Trained machine learning models to classify if emails are spam or not spam using **4600 emails** in Spambase data set
- Leveraged supervised algorithms **Decision Tree**, **K-Nearest Neighbor**, **Naive Bayes**, **SVM** & attained test accuracy of **92%**

US Election Data Exploration and Modelling | *Python, Sklearn, Matplotlib, Pandas*
- Performed data **preprocessing** & **Exploratory Data Analysis** on 2018 US Midterm Election Results & US Demographic Data
- Built **Regression**, **Classification** & **Clustering** models to **predict winning party** with a test accuracy of **85%**

EXPERIENCE

Undergraduate Research Assistant under Prof. Richard Joseph Jul 2018 – Apr 2019
- Architected **Azure ML** predictive model as an API to forecast drought-prone regions using weather data of the past **25 years**
- Achieved test accuracies of **92%** & **94%** on the dataset with **SVM** and two-class **decision tree** models respectively
- Received **AI for Earth** Azure Compute Grant worth $15,000 from **Microsoft** & **National Geographic**

Summer Project Trainee, Bhabha Atomic Research Centre, India May 2018 – Jul 2018
- Facilitated the optimization of a **data acquisition pipeline** for Low-Temperature Calorimetry experiments in 6 weeks
- Migrated LabWindows code into **LabVIEW** for nano voltmeter, milliammeter, and current source resulting in 70% performance improvement of data acquisition and increased numeric precision of experimental observations

Undergraduate Research Assistant under Prof. Dr. Mrs. Gresha Bhatia Aug 2017 – Mar 2018
- Designed a web app for users to monitor the **daily electricity consumption** of appliances & check against faulty power bills
- Awarded UGC **Minor Research Grant** by University of Mumbai under domains of Machine Learning & Internet of Things
- Published Springer paper **Interactive Electricity Consumption System** at SSIC 2019