MRINAL SONI

+16179496296 | soni.mr@northeastern.edu | LinkedIn | GitHub

EDUCATION

NORTHEASTERN UNIVERSITY – Boston, MA, USA *Sep 2018 – Dec 2020*
Master of Science in Data Science

SIR PADAMPAT SINGHANIA UNIVERSITY – Udaipur, Rajasthan, India *Sep 2013 – Apr 2017*
Bachelor of Technology in Computer Science Engineering

WORK EXPERIENCE

Northeastern University – Boston, MA
Graduate Teaching Assistant (Intermediate Programming with Data) *Oct 2020 – Dec 2020*
- Enhanced student learning by ensuring 1-on-1 interaction & assistance with concepts of Python & Data Science
- Collaborated in developing course structure, grading & creating rubrics for fellow TAs accelerating grading process by 35%
- Conducted lab sessions & office hours consisting of 30 students clarifying doubts with programming, providing feedback leading to efficient submissions & an increase of 9 points on average for new submissions

Northeastern University – Boston, MA
Graduate Teaching Assistant (Database Management Systems) *Jan 2020 – Apr 2020*
- Assisted 10 students with programming in SQL, key concepts of Database Management Systems like Triggers, Stored Procedures, & Relational Algebra & setting up MySQL & Oracle

Spundhan – Belgaum, India
Data Analyst Intern *May 2017 – Nov 2017*
- Designed & launched a data staging pipeline for 5 clients introducing a centralized storage deploying relational database in MySQL eliminating manual labor of data entry
- Leveraged Talend & it's ETL services to integrate data from Excel to MySQL
- Automated ETL process by writing stored procedures & triggers boosting run-time performance by 60%
- Performed A/B testing periodically for internet companies analyzing customer behavior & increasing sales by INR 15,000
- Identified, analyzed, interpreted trends or patterns in complex datasets & visualized company-wide valuable insights in Tableau for marketing team, stakeholders & executives increasing sales by approximately INR 50,000

PROJECTS

OkCupid Profile Recommendation *Feb 2021 – Present*
- Designed a recommendation model utilizing topic modelling, clustering, & rule-based filtering for California users generating most ideal profile recommendations from a pool of 50,000 profiles eliminating hours of unnecessary swiping
- Illustrated distribution of users in California visualizing insights into demographic information using Plotly for 50,000 profiles

Cardiovascular Disease Prediction *Sep 2020 – Dec 2020*
- Developed 6 classification models to predict presence of Cardiovascular Disease to promote early detection / treatment
- Achieved a minimum of 80% accuracy for built models with a maximum of 0.76 recall value
- Established new statistical & psychological features by feature engineering through design & obtaining patient segmentation deriving valuable insights into what factors contribute to cardiovascular disease

Regenerating images & videos in TensorFlow *Sep 2019 – Dec 2019*
- Researched & built a SRGAN model to enhance low resolution images by applying deep learning with adversarial network to generate high resolution images
- Down sampled images using LANCZOS algorithm generating low-resolution images
- Trained the model over multiple epochs on the Indian Movie Face Database (IMFDB), a large unconstrained database containing 34k images leading to a more diverse database
- Tested the model on video frames extracted from ChoKePoint dataset (48 videos - 62,204 face images) using AMI in an EC2 instance successfully reconstructing videos of high quality

Topic Modelling using LDA & NMF on medical transcriptions *Sep 2019 – Dec 2019*
- Implemented topic modelling on medical transcriptions summarizing top 15 topics describing disease, diagnosis & treatment of a patient saving significant amount of time during consultation for future revisits
- Pre-processed raw corpus by cleaning, normalizing, lemmatizing, tokenizing healthcare data with NLTK
- Used doc2bow, TF-IDF & n-grams to generate tokens for LDA & Non-Negative Matrix Factorization
- Increased coherence scores from 0.67 to 0.78 for NMF & 0.51 to 0.67 for LDA tuning hyperparameters
- Visualized top 15 topics obtained by LDA using pyLDAvis for effective data analysis understanding gist of transcriptions

TECHNICAL SKILLS

Languages:	Python, R, Java, SQL
Packages:	Pandas, NumPy, SciPy, Scikit-Learn, TensorFlow, Keras, Matplotlib, NLTK, PyTorch, OpenCV, PIL, spaCy, Plotly
Tools:	IntelliJ, RStudio, Jupyter Notebook, Tableau, Talend
Databases:	MySQL, Oracle, PostgreSQL
Cloud:	Amazon Web Services (EC2, S3)
Machine Learning:	Linear Regression, Clustering, Classification, Topic Modeling, Dimensionality Reduction, Statistical Inference, Natural Language Processing, Recommender Systems, Hypothesis Testing (A/B Testing)