Sharvi Tomar

About me ☎908-417-3435 sharvi-tomar ✉stomar2@illinois.edu

Education

- **University of Illinois Urbana-Champaign** — Dec '22 (Expected graduation)
  M.S. in Statistics (focus on Data Science and Machine Learning) — *GPA: 4.0/4.0*
- **Delhi Technological University** — Aug '15 - May '19
  B.Tech in Mathematics & Computing — *Top 10% in class*

Skills

Languages: R, Python, SQL, MATLAB, SPSS, Java, JavaScript, ReactJS
Modules: NumPy, SciPy, Pandas, scikit-learn, matplotlib, TensorFlow, Keras, Pytorch, Git
Skills: Statistical Modeling, Regression Analysis, Predictive Modeling, Natural Language Processing, Visualization, Operations Research, Probability Distributions, Hypothesis Testing, Computer Vision, Vision transformers

Experience

University of Illinois Urbana-Champaign / *Graduate Teaching Assistant* — Jan '21 - Present

- Leading lab sessions, assisting lectures, holding weekly office hours, preparing answer keys, grading programming assignments and communicating project proposals of the course *IS407 Introduction to Data Science*

HSBC / *Software Engineer* — Jul '19 - Jul -'21

- Built Deep Learning (BERT) based Named Entity Recognition model to perform auto-indexing of SWIFT messages in correct fields to reduce manual handling of 280 messages/day
- Developed web-app using spring boot microservices & ReactJS to migrate task scheduling & report generation app
- Modelled regression pack using Tosca for app testing & deployed with Jenkins; reducing 3 work hours/release
- Mentored 3 engineers to streamline deployments, environment setup & resources

PricewaterhouseCoopers (PwC) / *Data Science Intern* — Jun '18 - Jul '18

- Engineered incidents classification system using Natural Language Processing; reduced response time by 22%
- Created clients' performance analytics dashboard on various KPIs & integrated with on-premise ERP system

Projects & Publications

Implemented multi-layer neural networks with Stochastic Gradient Descent & Adam optimizer from scratch [Kaggle]

- Coded forward & backpropagation; achieved accuracy of 87.94% on Fashion-MNIST dataset with 3-layer network

Implemented Logistic Regression, Perceptron, SVM, Softmax classifier from scratch [Kaggle]

- Programmed linear classifiers for binary & multi-class image classification

Classification of Hand Gestures in Sign Language to English letters [Report] [Slides]

- Designed bucketing & ensembling approach; improved model's accuracy from 84.3% with single classifier to 86.4%
- Achieved on avg. 7.9 % reduction in misclassification rate per label for similar hand gesture signs

Real Estate Valuation Prediction [Report]

- Predicted Per Unit Area of the houses with best RMSE= 6.9 using Random Forest model
- Compared various machine learning algorithms & identified most influential factors for real estate valuation

Insights to modeling COVID-19 using new labels. *Towards Data Science*, May, 2020. [Article] [Code]; *2k+ views*

Categorization of Short-text based on Context & Semantics [Report]

- Performed semantic categorization by N-gram model & topic modeling using LDA, Dirichlet Multinomial Mixture model on Quora Questions
- Used K-means clustering on vectorized topics & assigned cluster centroids as superclass of generated topics
- Attained Perplexity value of 140 on test data & best KL Divergence score of 1.5

Genre-based Lyrics Generator & Evaluator [Report] [Slides]

- Built NLP model - three-layer character-level LSTM-RNN network to generate new lyrics based on song genres
- Achieved a cosine similarity index of 0.45 on 5000 song lyrics from 4 different genres

Best Smartphone Selection using TOPSIS

- Devised an innovative approach based on optimizing algorithms to solve multi-criteria decision problem
- Implemented enabling feature for users to set weights to each criterion to obtain best-suited phones analytically

A Survey on Internet Traffic Classification Techniques. *International Journal of Computer Engineering and Applications, Volume XII, August 2018, www.ijcea.com ISSN 2321-3469.* [Link]