Masood Dastan

Data Scientist, Economist
(470) 775.6115 · masooddastan@gmail.com · **Portfolio:** masooddastan.github.io

As a passionate data scientist with a strong foundation in economics, I thrive on the convergence of mathematics, statistics, and economic principles to uncover valuable insights. My journey blends academic expertise in teaching and research with a keen interest in unraveling complex challenges through data-driven methodologies. Committed to continuous growth and innovation, I leverage advanced data analytics and deep learning techniques to drive positive change and make a meaningful impact.

EXPERIENCE

Data Science Fellow, General Assembly | Remote | May 2023 – Aug 2023

- 400+ hour training program focusing on developing skills to analyze, interpret, and effectively communicate data-driven insights from massive data sets, and to predict what happens next through predictive modeling and pattern recognition.
- Demonstrated hands-on experience by successfully executing various data science projects. Utilized a wide array of techniques and tools including data cleaning, data manipulation and extraction using SQL, data visualization, feature engineering, and statistical modeling techniques such as linear regression and regularization.
- Expertise extended to machine learning methodologies like cross-validation, decision trees, ensemble models, and support vector machines. Proficiency was also gained in advanced areas such as Natural Language Processing (NLP) and text analysis.
- Acquired substantial knowledge in deep learning, encompassing neural networks such as Deep Neural Networks (DNN), Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN), among others.

Assistant Professor, The University of Texas at El Paso | TX, USA | Sep 2020 - present

- Instruct and develop undergraduate and graduate courses in machine learning, business analytics, and economics.
- Designed and delivered a comprehensive curriculum that covered a wide array of topics, including classification and regression, hypothesis testing, causal inference, Time series analysis and forecasting, data cleaning, exploratory data analysis (EDA), best practices in data visualization, and data querying using SQL.
- Authored and published a peer-reviewed paper applying advanced statistical techniques to address missing data in assessing corruption's impact on firm investment in developing countries; emphasized findings in international conferences.

Visiting Assistant Professor, Texas A&M International University | TX, USA | Sep 2019 – Aug 2020

- Instructed courses in Econometrics, Microeconomics, and Macroeconomics.
- Delivered lessons on econometric modeling, encompassing topics such as ordinary least squares, instrumental variable estimation, limited dependent variable models, time series models, and panel data methods.

EDUCATION

Certificate - Data Science Immersive | **General Assembly** | Aug 2023
Ph.D. in Economics | **Georgia State University** | Georgia, USA | Aug 2019
M.A. in Energy Economics | **University of Tehran** | Tehran, Iran | May 2013
B.A. in Economics | **University of Tehran** | Tehran, Iran | May 2010

CERTIFICATIONS

Python for Data Science and Machine Learning Bootcamp, Udemy
The Ultimate MySQL Bootcamp: Go from SQL Beginner to Expert, Udemy
Google Data Analytics, Coursera

SKILLS

Data Analysis Skills: Machine Learning, Econometrics, Time Series Analysis, Causal Analysis, Deep Learning, NLP
Language Proficiencies: Python, SQL, R, Tableau, STATA
Other Misc. Technologies: Git, AWS, Spark, LaTeX

DATA SCIENCE PROJECTS

Investigating the Impact of Federal Open Market Committee (FOMC) Meetings on Stock Market Performance | SpaCy, NLTK, Scikit-Learn, Transfer Learning, Deep Learning

- Leveraged advanced NLP techniques with SpaCy, NLTK, Scikit-Learn, and deep learning models to analyze and extract invaluable insights from FOMC meeting minutes.
- Employed topic modeling (LDA) to discern significant US economic concerns deliberated during FOMC meetings.
- Executed sentiment analysis using zero-shot text classification to assess FOMC meetings.

Reddit Post Classification | Python, NLTK, Scikit-Learn

- Employed Reddit API to gather 10,000+ posts from "Personal Finance" and "Investing" subreddits.
- Utilized NLTK, CountVectorizer, TF-IDF, and Scikit-Learn models for data preprocessing and data modeling.
- Implemented hyperparameter tuning via randomized search, encompassing a spectrum of classification methods like Naïve Bayes, Logistic Regression, Random Forest, Support Vector Machine, and Gradient Boosting.
- Achieved exceptional model performance, surpassing benchmark accuracy rates by an impressive 30 percentage points and attaining a remarkable 97% AUC score.

Predicting Credit Card Delinquency | Python, Scikit-Learn, Deep Learning, Gradient Boosting

- Led a collaborative team effort to create a predictive model for credit card delinquency, ensuring effective version control and seamless collaboration using Git and GitHub.
- Spearheaded data cleaning, exploratory data analysis (EDA), and feature engineering to optimize data quality.
- Implemented diverse machine learning models for credit delinquency prediction, culminating in a finely-tuned Adaboost model with an 80% accuracy in the test data—signifying a remarkable 27 percentage point improvement over the benchmark.

Housing Market Analysis | Python, Scikit-Learn

- Enhanced data quality by rectifying missing values, outliers, and incorrect data types.
- Utilized linear regression, Lasso, Ridge, and Elastic Net models. Implemented forward selection to identify impactful variables, resulting in a model surpassing the benchmark by 75%.