Tural Sadigov

U.S. Permanent Resident

December 2022

📍 Hamilton College, 198 College Hill Rd, Clinton, NY

📞 +1 (812) 650 - 2583

🏠 turalsadigov.github.io/

✉ tsadigov@hamilton.edu

○ turalsadigov

in tural-sadigov

About me ─────

Visiting Assistant Professor of Mathematics and Statistics with Ph.D. in Applied Mathematics and 7+ years experience in developing and teaching both mathematics and data-related courses such as various levels of Statistics, Machine Learning, Probability, and Time Series Analysis (on Coursera) and mentoring undergraduate Machine Learning projects in Statistical Methods in Machine Learning course. Using R and RStudio extensively in all data-related courses. Creator of R package stats2data for the Statistical Modeling and Applications course at Hamilton College. Problem solver since his high school Mathematical Olympiad years. More detailed cv: `https://turalsadigov.github.io/cv.html`

Professional Summary

7 + years of experience in Applied Mathematics, Statistics, Machine Learning and Data Science.

- Classical Statistics
  - Expert at Descriptive Statistics, Inferential Statistics, Hypothesis Testing, Resampling methods, Bootstrap methods, Permutation Tests. Equipped with experience in utilizing statistical techniques which include Correlation, Hypotheses testing, modeling, Inferential Statistics as well as descriptive statistics techniques.
  - Expert at working with Statistical Tests: two-way independent & paired t-test, Chi-squared tests, Permutation Tests, etc.

- Machine Learning and modeling
  - Experienced in structured data analytics and developing Machine Learning algorithms to gain operational insights. Proficient in Tree Based Methods and Ensemble Learning using Decision Trees, Bagging, Boosting & Random Forests.
  - Expertise in building Supervised and Unsupervised Machine Learning experiments using R performing detailed predictive analytics for all types of data: continuous numerical variables, discrete numerical variables, nominal categorical variables, and ordinal categorical variables.
  - Well versed with Linear/non-linear, regression and classification modeling predictive algorithms. Actively involved in model fitting, model validation, hyper-parameter tuning and model selection
  - Expertise in using Multiple Regularized Linear & Logistic Regression and Classification Modeling, Decision-trees, PCA, K-Means, K-NN algorithms and have authored and coauthored several scholarly articles on applied mathematics.
  - Used PCA, a Dimensionality Reduction technique, and K-means clustering for understanding of high-dimensional data sets
  - Evaluated the classification algorithms using Precision, Recall, F1-score, Confusion Matrix, Area under the Curve (AUC) for Precision-Recall Curve and ROC Curves
  - Teaching level proficiency in Time Series Analysis: particularly, ARIMA and SARIMA models, and forecasting techniques such as Simple, Double and Triple Exponential Smoothing
  - Familiar with tensorflow and keras (Deep Learning) both in Python and R

- ETL
  - Extensively involved in Data Wrangling, Cleaning, Preprocessing & Feature Engineering, Data preparation, Exploratory analysis, feature engineering using Supervised and Unsupervised modeling. Handled imbalanced datasets, exploring the uses of under-sampling, over-sampling, and SMOTE.
  - Performed preliminary data analysis using descriptive statistics, introducing dummy (indicator) variables, and handled anomalies such as removing duplicates and imputing missing values using various imputation method such as mean, median and KNN imputations.
  - Extensive experience in Data Cleaning, Feature Engineering, Machine Learning and Data Science technologies ensuring student project completion on time and with the desired results. Performed data cleaning, features scaling, features engineering using R, tidyverse and tidymodels in R Studio
  - Performed feature scaling, feature engineering and statistical modeling. Extensive modeling experience, particularly within R Studio

- Visualization and communication

  - Strong experience and knowledge in data visualization with ggplot2 (tidyverse) creating Scatterplots, Dotplots, Histograms, Smoothed Histograms, Bar plots, Pie-charts, Boxplots, Time plots, Violin plots etc. Performed Exploratory Data Analysis with R and Visualization tools like ggplot to identify the patterns and correlations between the features.

  - Communication, reporting, visualization and dashboard: Quarto, R Markdown, Jupyter Notebook, R Shiny.

  - Sample R Shiny web app: `https://turalsadigov.shinyapps.io/my_new_ames_app/`.

- Leadership and collaboration

  - Actively advising and mentoring students in data science and mathematics majors at Hamilton College, and supervised all adjunct lecturers at SUNY Polytechnic Institute as Math Service Coordinator (position similar to an associate chair)

  - Strong decision-making ability with the aid of data analysis, and expert judgments. A quick and a life-long learner as evident from career/area change (from Partial Differential Equations to Data Science) right after PhD.

  - Team player with good logical reasoning ability, coordination and interpersonal skills. Team builder with excellent communications, time & resource management

  - Collaborated on Machine Learning projects with interdisciplinary group of students coming from all different majors, and worked on applied mathematics research with various teams of mathematicians and scientists across the nation

Education

2008 - 15	Ph.D and MA, Applied Mathematics Bloomington, IN	Indiana University
2003 - 08	BS, Mathematics Istanbul, Turkey	Boğaziçi University

Technical Expertise, Data Science and Coding Skills

- R: Tidyverse, Tidymodels, Infer, ggplot2, parsnip, yardstick, rsample, recipe, dplyr, tidyr, base R

- Python: Pandas, Numpy, Scikit-learn, TensorFlow, Keras, Matplotlib, Seaborn

- Regression: Linear, Multiple Linear, Polynomial, Nonlinear (through variable transformation), Regularized Regression (LASSO, l2-regularization), KNN regression, Regression Trees, Bagged Models, Random Forest, Fully Connected Neural Networks, Principal Component Regression

- Classification: Logistic Regression, Multiple Logistic Regression, Regularized Logistic Regression, KNN classification, Bayes Classifier, Classification Trees, Random Forest, Bagged Models, Maximal Margin Classifier, Support Vector Classifier, Support Vector Machine, Fully Connected Neural Networks

- Unsupervised Learning: K-means clustering, Hierarchical clustering, Principal Component Analysis, Singular Value Decomposition

- SQL

Professional Experiences: Current and Past Positions

- Visiting Assistant Professor, 2020 - current, Hamilton College

  - Redesigned a series of data science courses such as

    * Statistical Analysis of Data

    * Statistical Modeling, and Its Applications

    * Statistical Methods in Machine Learning

  - Incorporated heavy R coding into lectures, assessments, and projects

- Actively participated in data science initiatives of Hamilton College where one such initiative resulted in a data science major
- Successfully engaged students in class Machine Learning projects where designed and developed state-of-the-art predictive Classification and Regression models using Machine Learning – Supervised and Unsupervised algorithms and performed in-depth analysis on the structure of models.
- Extensively used open-source tools – Python and R Studio (R) for statistical analysis and building machine learning algorithms.
- Applied concepts of R-squared, Adjusted R-squared, MSE, RMSE, MAE, p-values, residual analysis, validation set approach and cross-validation techniques in the evaluation stage to extract interesting findings through comparisons.
- Played a crucial role in increasing the number of majors in mathematics and data science
- Taught Calculus courses with focus of applying mathematics

- Coursera Instructor , 2017 - current, Coursera

- Developed and taught Practical Time Series Analysis (modeling and forecasting) with a colleague
- Served 72988 students as of November, 2022

- Applied Mathematics Lecturer and Math Service Coordinator, "2015 - 20, SUNY Polytechnic Institute
  - Promoted data analysis certificate
  - Delivered technical lectures in Applied Probability
  - Delivered technical lectures in Regression
  - Delivered technical lectures in Time Series Analysis
  - Delivered technical lectures in math courses such as various levels of Calculus, Linear Algebra and Differential Equations.
  - Delivered technical lectures in undergraduate and graduate levels

Selected mentored ML projects (All are teamwork)

Also see https://github.com/turalsadigov/mentored_projects.

1. *Does Economic Development Predict Democratization?*, With Chiara Bondi, John Madigan, 2022

   - Methods: Joining dataset from various sources, data wrangling (pivoting), training and testing split, data imputation, feature engineering, Regularized Regression Models (LASSO, Ridge, Elastic-Net), cross-validation and hyper-parameter tuning
   - Packages: tidyverse, caret, glmnet

2. *A Regularized Binomial Logistic Regression Approach to Cancer Classification Using Gene Expression*, Joshua Horowitz, Mukund Jayaran, 2022

   - Methods: feature engineering, training and testing split, Multiple Logistic Regression with L1 regularization,
   - Packages: tidyverse, caret, glmnet

3. Predicting Career Longevity of NBA Rookies, Margaret Phipps, Luke Devine, 2021

   - Methods: feature engineering, training and testing split, Support Vector Classifier, Support Vector Machines (with polynomial and radial kernels), cross validation and hyper-parameter tuning, confusion matrices
   - Packages: caret, e1071, kernlab

4. Identifying Parkinson's Disease Through Speech Patterns, Ian Nduhiu, Lindsay Gearty (Methods: Support Vector Machines), 2021

   - Methods: training and testing split, Support Vector Classifier, Support Vector Machines (with polynomial and radial kernels), cross validation and hyper-parameter tuning, confusion matrices, ROC curves, AUC comparison
   - Packages: caret, e1071, kernlab, roc

5. Predicting Housing Rent Prices Using House Characteristics, Taron Kui, Iftikhar Ramnandan, Jenny Tran, 2021
    - Methods: examination of multicollinearity, training and testing split, pruned regression/decision tree, bagged model and random forest, cross validation and hyper-parameter tuning
    - Packages: rpart, tree, tidyverse

6. How Race, Income, and Education Relate to Internet Access in US Counties, Lindsay Gearty, Margraet Phipps (Methods: Multiple Polynomial Regression, Forward/Backward Variable Selection), 2021
    - Methods: feature engineering, multiple linear regression, polynomial regression, backward variable selection using various measures, examination of multicollinearity
    - Packages: MASS, glmnet

Research Experience

2020, 21	Summer Research Associate Air Force Research Lab Griffis Institute

Proved the existence of theoretical neural network solutions for differential equations, and designed and implemented numerical algorithms in Python to solve partial differential equations. This was a teamwork consisting of two research scientists, me, and one of my students.

Awards/Grants

- Summer research grants, Air Force Research Lab/Griffiss Institute, 2021-22 ($36000)
- Dean's Pedagogical Development Awards, Hamilton College, 2021-22 ($4500)
- Teaching Grant from Herkimer College and Hamilton College for teaching Financial Mathematics, 2022 ($10000)
- SGU Award for Excellence in Teaching, SUNY Poly, AY 2018-19
- Bronze Medal, 44th International Mathematical Olympiad, Tokyo, Japan, 2003
- Gold Medal (four times), Azerbaijan Mathematical Olympiad, 2000-03

Selected data-related talks

1. Theoretical Analysis of Artificial Neural Network solutions to the damped, driven Korteweg de-Vries equation, AMS Special Session, 2022

2. Support Vector Machines, CUNY - Hunter College, 2021

3. Model Selection & Validation in Machine Learning, CUNY - Hunter College, 2021

4. Support Vector Machines: Overview and Applications, CUNY - Hunter College, 2020

5. Support Vector Machines: Overview and Applications, Hamilton College, 2020

6. Stochastic Calculus: Ito Integrals and Stochastic Differential Equations, SUNY Poly, 2019

7. Stochastic Processes, Markov Chain and Brownian Motion, SUNY Poly, 2019

Publications

1. Safety Prediction Model for Reinforced Highway Slope using a Machine Learning Method, 2020 (Link to the paper)
2. A Determining Form for the Subcritical Surface Quasi-Geostrophic Equation, 2019 (Link to the paper)
3. Determining form and data assimilation algorithm for weakly damped and driven Korteweg–de Vries equation — Fourier modes case, 2017 (Link to the paper)
4. A determining form for the damped driven nonlinear Schrödinger equation—Fourier modes case, 2015 (Link to the paper)