Tural Sadigov

U.S. Permanent Resident

November 2022

Hamilton College, 198 College Hill Rd, Clinton, NY

+1 (812) 650 - 2583

turalsadigov.github.io/

tsadigov@hamilton.edu

turalsadigov

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 applying Statistics based Data Analytics using R Machine Learning techniques.
- Expert at Descriptive Statistics, Inferential Statistics, Hypothesis Testing, Resampling methods, Bootstrap methods, Permutation Tests
- Expert at Data Wrangling, Cleaning, Preprocessing & Feature Engineering, Data Visualization
- Experienced in structured data analytics and developing Machine Learning algorithms to gain operational insights
- Extensively involved in Data preparation, Exploratory analysis, Feature engineering using Supervised and Unsupervised modeling.
- Well versed with Linear/non-linear, regression and classification modeling predictive algorithms. Actively involved in Model selection, Statistical analysis, analyzing Correlations and similarities.
- Expert at working with Statistical Tests: two-way independent & paired t-test, one-way & two-way ANOVA along with Non-parametric tests, Chi-squared tests, Permutation Tests.
- Handled the imbalanced dataset, exploring the uses of under-sampling, over-sampling, and SMOTE.
- Proficient in Tree Based Methods and Ensemble Learning using Decision Trees, Bagging, Boosting & Random Forests.
- Created dashboards as part of Data Visualization using ggplot2
- Communication, reporting and dashboard: Quarto, R Markdown, Jupyter Notebook, R Shiny. Sample R Shiny web app: Link to the app
- 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.
- 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.
- Teaching level expertise and proficiency in Time Series Analysis: particularly, ARIMA and SARIMA models
- Extensive modeling experience, particularly within R Studio
- Familiar with tensorflow and keras (Deep Learning) both in Python and R
- Leadership: have been advising many students in math and data science majors at the college, was the Math Service Coordinator in my previous job
- Strong decision-making ability with the aid of data analysis, and expert judgments. A quick and a Life-long learner.
- Team player with good logical reasoning ability, coordination and interpersonal skills.
- Team builder with excellent communications, time & resource management

Education

2008 - 15 Ph.D and MA, Applied Mathematics — Indiana University
Bloomington, IN

2003 - 08 BS, Mathematics — Boğaziçi University
Istanbul, Turkey

Technical Expertise, Data Science and Coding Skills

- Python: Pandas, Numpy, Sklearn, TensorFlow, Keras, Matplotlib
- R: Tidyverse, Tidymodels, Infer, ggplot2, parsnip, recipe, dplyr, tidyr, base R
- 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
- 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
- SQL, dplyr

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.
  - Skilled in Advanced Regression Modelling, Correlation, Multivariate Analysis, Model Building, and application of Statistical Concepts.
  - 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 Exploratory Data Analysis with R and Visualization tools like ggplot to identify the patterns and correlations between the features.
  - Used PCA, a Dimensionality Reduction technique, and K-means clustering for understanding of high-dimensional data sets
  - Performed data cleaning, features scaling, features engineering using R and tidyverse in R Studio
  - Analytic, performance-focused, and detail-oriented professional, offering in-depth knowledge of data analysis and statistics.
  - Equipped with experience in utilizing statistical techniques which include Correlation, Hypotheses testing, modeling, Inferential Statistics as well as descriptive statistics techniques.
  - Expertise in building Supervised and Unsupervised Machine Learning experiments using R and Python performing detailed predictive analytics for all types of data: continuous numerical variables, discrete numerical variables, nominal categorical variables, and ordinal categorical variables.

  - 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.
  - Evaluated the classification algorithms using Precision, Recall, F1-score, Confusion Matrix, Area under the Curve (AUC) for Precision-Recall Curve and ROC Curve.
  - Mitigated risk factors through careful analysis of financial and statistical data. Transformed and processed raw data for further analysis, visualization, and modeling.
  - Performed feature scaling, feature engineering and statistical modeling.
  - 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 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
  - Both in undergraduate and graduate levels

Selected mentored ML projects (All are teamwork)

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 Inter- net 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