Mustafa Murat ARAT Statistician | Data Scientist | Machine Learning (Deep Learning) Specialist

arat.murat@gmail.com | Ankara, Turkey | www.mmuratarat.github.io
github.com/mmuratarat | linkedin.com/in/mmuratarat

SUMMARY

PhD-level educated, bilingual statistician and data scientist, who works to design, develop and implement advanced data-driven predictive methods, powered by machine learning and deep learning algorithms, to provide actionable insights from large volumes of real-world, structured and unstructured data in order to satisfy the business needs / goals for decision making across different industries and roles. I have a deep understanding of best data analytic practices to solve complex problems and I continuously improve myself and learn new emerging technologies, tools, and platforms to add new skills to my skill set.

SKILLS

Programming Languages: Python
Deep Learning Libraries: Tensorflow, and Keras.
Container and Orchestration Systems: Docker, Swarm, Kubernetes.
Database Systems: PostgreSQL (with pgadmin4 or DataGrip)
No-SQL/Full-text Search Systems: MongoDB and Elasticsearch/Kibana
Big Data Frameworks: Apache Spark
Scripting Languages: R, and MATLAB.
Visualization Software: Tableau.
Statistical Softwares: SPSS, and Minitab
Version Control: Git/Github
Workflow Management Systems: Metaflow and Airflow (with its GUI).
Markup Languages: HTML, LATEX, Beamer, Jekyll, and Markdown/R-Markdown.
Miscellaneous: Extensive usage of JupyterLab (notebooks) and comfortable with Bash command line interface.
Languages: Turkish (Native Language), English (Full professional proficiency), French (Elementary proficiency)

PROJECTS AND TRAININGS

- **Sales Elasticity of Emotional Displays**
  - This project was part of one of my chapters of my doctoral thesis through an American direct-to-consumer retailer company, that took place between years 2019 – 2020.
  - It marks the first empirical approach which establishes the sales elasticity of emotional displays in sales pitches.
  - We apply state-of-the art artificial intelligence technology to almost two years of video footage (17,312 hours or 62.32 million frames), in which a TV host makes her/his sales pitch to customers, to detect human faces and Salesperson's facial emotions are then extracted and classified into six emotional displays: happiness, sadness, surprise, anger, fear, and disgust. We, then, matched them to customers' purchase data to understand the true nature of the relationship. To estimate sales elasticity, we incorporate emotional displays in sales response models together with marketing mix activities (i.e., product, price, display duration, and free shipping).
  - We found that the sales elasticities of a host's emotional displays are uniformly negative, including that of happiness, which is a provocative finding (against "service with a smile" policy) because it partially contradicts the external validity of social contagion theory. In other words, emotional displays in sales pitches are hazardous to business, salespersons should pitch with neutral expressions (i.e., absence of emotion). Additionally, we found that the presence (versus absence) of a host's face in the frame positively increases sales by 0.62%. Therefore, the study offers guidance to firms on re-training low performers on "selling with a straight face" as well as to reward successful hosts
  - It was published in Journal of Marketing in February, 2022.

- **Intervention 10: Research on Occupational Accident Prediction Model**
  - Consultant / Senior Non-Key Expert for DAI Global, July 2022 – August 2023.
  - Part of the European Union Project "Technical Assistance for Strengthening Training and Research Capacity of the Centre for Labour and Social Training and Research (ÇASGEM)" (TREESP1.3.CASGEM/P-01, EuropeAid/140073/IH/SER/TR) by *Directorate of European Union and Finance Assistance (DEUFA)* in the *Ministry of Labour and Social Security* of the *Republic of Turkey*.
  - Having access to Turkish occupational accident database maintained by the Social Security Institution, between the years 2013-2021, we tried to predict the outcome of the incident for construction, mining, chemical, and metal sector separately.
  - 4 different multivariate regression models are developed to identify the potential risk factors influencing the outcome of an occupational accident. After the models are obtained, odds ratios have been computed for the significant variables and interpreted.
  - Due to the insufficient data points and class imbalance, modelling could not be performed in petrochemistry industry.
  - Final model for metal sector, influencing the outcome of an occupational accident, includes age, deviation group, OSH training, and medical intervention factors. AUC metric for this model is to be found 0.81.

- For construction sector, factors affecting the fatality of post-accident risk were identified as the number of insured workers, daily average income, age, duration of insurance, level of education (in years), hours the worker has been on duty (hours worked), the year of the accident occurrence, type of workplace, work environment, deviation group, medical intervention status, OSH training, having prior accidents within the same sector, and having prior accidents in a different sector. AUC metric for this model is to be found 0.84.
- For mining sector, significant factors were found to include age, level of education (in years), hours the worker has been on duty (hours worked), daily average income, deviation group, medical intervention status, OHS training, and whether or not there has been a previous accident in the same sector. AUC metric for this model is to be found 0.94.
- This intervention also acts as a vessel for policy-makers not just to support the enforcement of the relevant legislation but also to improve the collected data. It highlights the problems existing in the data and database about occupational accidents with possible solutions, such as increasing data quality, developing well-defined data governance policies and procedures and data validation tools (i.e., Great Expectations and AWS's Deequ).

- **Intervention 5.5.: Training on Advanced Statistical Analysis Methods for Health Physical and Social Sciences in Occupational Health and Safety**
  - Trainer / Senior Non-Key Expert for DAI Global, March 2022 - June 2022.
  - Part of the European Union Project "Technical Assistance for Strengthening Training and Research Capacity of the Centre for Labour and Social Training and Research (ÇASGEM)" (TREESP1.3.CASGEM/P-01, EuropeAid/140073/IH/SER/TR) by *Directorate of European Union and Finance Assistance (DEUFA)* in the *Ministry of Labour and Social Security* of the *Republic of Turkey*. www.casgemeuproject.org.
  - for 1st, 2nd and 3rd Sessions, Total 75 participants.
  - Covering multiple topics including but are not limited to analysis of variance (ANOVA) and analysis of variance (ANCOVA), linear regression, logistic regression, regularization techniques, generalized linear models, dimension reduction techniques, non-parametric methods, imbalanced learning and performing all these techniques using real-world datasets by using Python programming language and its libraries.

- **Intervention 5.4.: Training on Basic Statistical Analysis Methods for Health Physical and Social Sciences in Occupational Health and Safety**
  - Trainer / Senior Non-Key Expert for DAI Global, August 2021 - December 2021.
  - Part of the European Union Project "Technical Assistance for Strengthening Training and Research Capacity of the Centre for Labour and Social Training and Research (ÇASGEM)" (TREESP1.3.CASGEM/P-01, EuropeAid/140073/IH/SER/TR) by *Directorate of European Union and Finance Assistance (DEUFA)* in the *Ministry of Labour and Social Security* of the *Republic of Turkey*. www.casgemeuproject.org.
  - for 2nd and 3rd Sessions, Total 50 participants.
  - Covering multiple topics including but are not limited to measures of location and dispersion and their appropriate uses, random variables, discrete/continuous probability distributions, hypothesis testing, confidence intervals, Chi-square and tests of contingency tables, simple linear regression and one-way analysis of variance (ANOVA) and applying everything by using Python programming language and its libraries.

- **Training for "Machine Learning" |** *Presidency of Turkey, Presidency of Strategy and Budget*
  - Trainer, March 7, 2022 - April 1, 2022, 35 participants.
  - Covering multiple topics including but are not limited to data engineering (scaling the data, missing value imputation, outlier detection, encoding categorical variables, feature selection, hyper-parameter tuning, model evaluation criteria etc.), classification, clustering and regression algorithms, dimensionality reduction methods and performing these techniques using real-world datasets by using Python programming language.

- **Training for "Data Analysis with Python" |** *Presidency of Turkey, Presidency of Strategy and Budget*
  - Trainer, January 24, 2022 - February 7, 2022, 35 participants.
  - Covering multiple topics including but are not limited to primitive/non-primitive data structures in Python, control-flow, object-oriented programming, NumPy and pandas, string manipulation, data visualization etc.

- Trainer for 4 weeks online training for **"Artificial Intelligence"**, between March 26, 2022 and April 17, 2022, covering multiple topics Python, Statistics, Machine Learning / Deep Learning algorithms, around 70 participants, www.yapayzeka.us,

- **Two separate trainings for "Basic Statistical Methods" and "Advanced Statistical Methods" |** *HALKBANK Board of Inspection*
  - Trainer, November 7, 2022 – January 30, 2023, 6 hours per week, 30 participants.
  - Covering multiple topics including but are not limited to probability distributions, confidence intervals, hypothesis testing, data and feature engineering, linear regression, logistic regression, generalized linear models, regularization techniques, imbalanced classification approaches, (M)ANOVA methods and performing these techniques using real-world datasets by using Python programming language.

EDITOR AND TRANSLATOR

- Editor of the Turkish translation of the book "*Designing Machine Learning Systems*" by Chip Huyen.
- Editor of the Turkish translation of the book "*Automate the Boring Stuff with Python: Practical Programming for Total Beginners*" by Al Sweigart, August 2022.
- Editor of the Turkish translation of the book "*Python Crash Course: A Hands-On, Project-Based Introduction to Programming*" by Eric Matthes, March 2022.
- Editor of the Turkish translation of the book "*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems"* by Aurelien Geron, 2021.
- One of the translators of the Turkish translation of the book "*Schaum's Outline of Probability (2nd edition)*" by Marc Lipson and Seymour Lipschutz, 2013.
- One of the translators of the Turkish translation of the book "*Introduction to Probability*" by George G Roussas, 2014.

EDUCATION

UNIVERSITY OF TENNESSEE, KNOXVILLE Location (Knoxville, TN / USA)
Doctor of Philosophy in Analytics Years Attended (2015-2020)

- Recipient of Scholarship/Assistantship from Department of Business Analytics & Statistics of Tennessee, Knoxville for my doctoral studies between the years 2015 – 2020.
- Earned Statistics Excellence Fund from Department of Business Analytics & Statistics of Tennessee, Knoxville for the Fall semester of 2015.
- GPA: 3.92 / 4.00
- Coursework includes: Prescriptive Analytics, Statistical Inference, Probability/Stochastic Process, Bayesian Modeling and Computations, Data Mining Methods and Business Applications, Applied Multivariate Methods, Image Reconstruction, Reinforcement Learning, Applied Time Series, Probability and Mathematical Statistics, Multivariate Data Mining Techniques.
- Completed the dissertation "*Advances and Applications in Deep Learning*" under the supervision of Dr. Michel BALLINGS.

HACETTEPE UNIVERSITY Location (Ankara / TURKEY)
Master of Science in Statistics Years Attended (2011-2014)

- GPA: 3.93 / 4.00 (ranked 1st of 15)
- Coursework includes: Linear Models, Time Series, Stochastic Processes, The Analysis of Contingency Tables, Simulation Techniques, Multivariate Statistical Methods, Survival Analysis.
- Completed the thesis "*A Study on Support Vector Machines*" under supervision of Prof. Dr. Turhan MENTES.

HACETTEPE UNIVERSITY Location (Ankara / TURKEY)
Bachelor of Science in Statistics Years Attended (2007-2011)

- Secured Supreme Success Award from Faculty of Science of Hacettepe University for the semester 2009–2010.
- GPA: 3.24 / 4.00 (top 10%, ranked 9th of 91)
- Coursework includes: categorical Data Analysis, Time Series, Multivariate Statistical Methods, Stochastic Processes, Operation Research, Regression Analysis, Mathematical Statistics, Nonparametric Statistical Methods, Sampling, Survey Design, Linear Algebra, Econometrics, Biostatistics.

PROFESSIONAL EXPERIENCES

- **Research Assistant**, Department of Statistics, Hacettepe University, Ankara, Turkey, October 2011 - present.
    - Teaching a wide range of undergraduate-level courses from Bayesian Statistics to Probability Theory.
    - Academic consultant for undergraduate students.
- **Graduate Teaching Assistant**, Department of Business Analytics & Statistics, University of Tennessee, Knoxville, U.S.A., August 2015 - July 2020.
    - Assisting students during the lab sessions and marking written reports of the assignments for courses such as Categorical Data Analysis, Customer Analytics, Applied Multivariate Methods, Applied Time series and many more.
    - Co-creating a Deep Learning class for the first time in the faculty with my advisor Dr. Michel Ballings, covering a wide range of topics, from Backpropagation to Convolutional Neural Networks, from Long Short-term Memory to Autoencoders.

INVITED TALKS / SEMINARS

- "Bringing Business, Statistical Methods and Transfer Learning Together: Sales Elasticity of Emotional Displays – Large Scale Evidence for Selling with a Straight Face", Department of Statistics, of Hacettepe University, December 29th 2020 (face-to-face).

- "A New Livestream Retail Analytics Framework to Assess the Sales Impact of Emotional Display", Department of Statistics, of Middle East Technical University, May 27th 2021 (online).
- "Introduction to Machine Learning and Basic Machine Learning Applications", Workshop for "Technical Assistance for Improving the Detection Capacity of Turkish Customs Enforcement" European Union Project, Abant, Bolu, Turkey, November 12nd 2021 (face-to-face).
- "Intern Career Talk", Department of Computer Science, Manisa Celal Bayar University, April 4th 2022 (online).
- "Machine Learning Operations (MLOPS)", Analitik Fokus, EnerjiSA, September 29th, 2022 (online).
- "How to become a good data scientist?", Google Developer Student Clubs, Ankara Yıldırım Beyazıt University, November 10th, 2022 (face-to-face).
- "How to become a good data scientist?", IEEE Computer Society, Manisa Celal Bayar University, May 30th, 2023 (online) - https://www.youtube.com/watch?v=JO9GcuXYOpE
- "Data Science and Digital Jobs", Workshop for "Improving Job and Vocational Counseling Services (IQJVC)" European Union Project of Turkish Employment Agency (İŞKUR), Gemlik, Bursa, Turkey, June 7th 2023 (face-to-face).

PEER-REVIEWED PUBLICATIONS

- Tan, O., Hizli Sayar, GH., Ünsalver, B.Ö., **Arat, M. M.** and Karamustafalioglu, O. (2014). "The correlations of nicotine addiction with the levels of impulsiveness, depression and anxiety in obsessive-compulsive patients". Journal of Dependence, 15(3), 124 - 133.
- Gogcegoz Gül, I., Eryilmaz, G., Hizli Sayar, G., Özten, E., **Arat, M.M.**, and Tarhan, N. (2014). "Evaluation of the Efficacy of The Continuation Electroconvulsive Therapy in Treatment-Resistant Schizophrenia". Revista de Psiquiatria Clínica, 41(4), 90 - 94
- Eryilmaz, G. Gogcegoz Gül, I., Yorbik, O., and **Arat, M.M.** (2014). "Evaluation of Clinical Response According to Plasma Paroxetine Level in Paroxetine-Responsive Major Depression". International Journal of Internal Medicine, 3(3), 39 - 42.
- Tan, O., Hizli Sayar, G., Ünsalver, B.Ö., **Arat, M. M.** and Karamustafalioglu, O. (2015). "Combining transcranial magnetic stimulation and cognitive-behavioral therapy in treatment resistant obsessive-compulsive disorder". Anatolian Journal of Psychiatry, 16(3), 180 - 188.
- **Arat, M.M.**, Aktas, S. (2016). "Generalized Maximum Entropy Approach to Unreplicated Factorial Experiments". Statistics and Its Interface, 9(3), 295 - 302.
- Bharadwaj, N., Ballings, M., Naik, P., Moore, G.M., **Arat, M.M.** (2022). "A New Livestream Retail Analytics Framework to Assess the Sales Impact of Emotional Displays". Journal of Marketing, 86(1), 27 – 47.
- **Arat, M.M.** (2022). "Learning From High-Cardinality Categorical Features in Deep Neural Networks". Journal of Advanced Research in Natural and Applied Sciences, 8 (2), 222 – 236.

ORAL PRESENTATIONS

- **Mustafa Murat Arat**, "Testing Export-Led Growth Hypothesis: The Case of Turkey, 1961-2010", Applied Statistics 2012, Ribno, Slovenia, September 23 – 26, 2012.
- Elcin Ergin, **Mustafa Murat Arat**, Cem Iyigun, Inci Batmaz, "Short-Term Electricity Load Forecasting Via Nonparametric Prediction Methods", EURO-INFORMS Joint International Meeting: 26th European Conference on Operational Research, Rome, Italy, July 1 - 4, 2013.
- **Mustafa Murat Arat**, Elcin Ergin, "Short Term Load Forecasting Using Support Vector Regression", European Conference on Data Analysis by The German Classification Society (GfKl) and the French speaking Classification Society (SFC), Luxembourg City, Luxembourg, July 1 – 4, 2013.
- **Mustafa Murat Arat**, Serpil Aktas Altunay, "Generalized Maximum Entropy Approach To Unreplicated Factorial Experiments", The 13th Annual Conference of the European Network for Business and Industrial Statistics (ENBIS-13), September 15 – 19, 2013.
- **Mustafa Murat Arat**, "Comparison of SVM and LS-SVM For Regression", y-BIS 2013: Joint Meeting of Young Business and Industrial Statisticians, sponsored by International Society for Business and Industrial Statistics (ISBIS) and European Network for Business and Industrial Statistics (ENBIS), Istanbul, Turkey, September 19 – 21, 2013
- Michel Ballings, Neeraj Bharadwaj, Prasad Naik, George Miller Moore, **Mustafa Murat Arat**, "But Wait, There's More! Deep Learning of Sales Elasticity of Sales Pitches", Theory + Practice in Marketing Conference, Columbia University, New York, NY, USA, May 18, 2019.

- **Mustafa Murat Arat**, George Miller Moore, Michel Ballings, "Maximizing Insights from Customer Data Streams", Interactive Marketing Research Conference, Houston, TX, USA, March 27 – 29, 2019.
- **Mustafa Murat Arat**, Michel Ballings, George Miller Moore, "Breaking Through Barriers to Deep Learning Adoption in Customer Behavior Modeling", INFORMS Annual Meeting 2019, Seattle, WA, USA, October 20 - 23, 2019.
- Neeraj Bharadwaj, Michel Ballings, Prasad Naik, Miller Moore, **Mustafa Murat Arat**, "Purchase Impact Of A Salesperson's Facial Expressions: Large-Scale Video Analysis Using Deep Learning", 2020 Winter AMA Academic Conference, San Diego, CA, USA, February 14 - 16, 2020.
- Osman Tolga Kaskati, **Mustafa Murat Arat**, Fatma Kaymakamtorunlari, Deniz, Emre Keskin, "Popülasyon Genetiği Çalışmalarında Mantel Testi Üzerinde Bir Uygulama", 22nd Local and 5th International Biostatistics Conference, Online, October 28 - 30, 2021.

Underline means who presents the paper.

COLLOQUIUMS/MEETINGS/CONFERENCES ATTENDEE

- 8th International Statistics Students Colloquium, Izmir, Turkey, May 14 - 15, 2011.
- The Joint Meeting of Young Business and Industrial Statisticians and Young Portuguese Statisticians, Lisbon, Portugal, July 23 - 26, 2012.
- 9th International Statistics Days Symposium, Antalya, Turkey, May 10 - 14, 2014.

SCIENTIFIC ACTIVITIES

- GIS 2013: Genç İstatistikçiler Sempozyumu (Young Statisticians Symposium), Member of the Local Organizing Committee, Ankara (Turkey), September 10 -11 2013, http://www.gis2013.hacettepe.edu.tr.
- AIK 2013: Araştırmacılar ve İstatistikçiler Konferansı (Researchers and Statisticians Conference), Member of the Local Organizing Committee, Ankara (Turkey), September 10 -11 2013, http://www.aik2013.hacettepe.edu.tr.
- y-BIS 2013: Joint Meeting of Young Business and Industrial Statisticians, sponsored by ENBIS (European Network for Business and Industrial Statistics) and ISBIS (International Society for Business and Industrial Statistics), Member of the International Scientific Committee and Organizing Committee, Istanbul (Turkey), September 19 – 21, 2013, http://ybis13.msgsu.edu.tr.
- y-BIS 2019: Joint Meeting of Young Business and Industrial Statisticians, sponsored by ENBIS (European Network for Business and Industrial Statistics) and ISBIS (International Society for Business and Industrial Statistics), Member of the International Scientific Committee and Organizing Committee, Istanbul (Turkey), September 25 – 28, 2019, http://ybis2019.msgsu.edu.tr.