MUSTAFA MURAT ARAT

Statistician | Data Scientist | Machine Learning (Deep Learning) Specialist

@ arat.murat@gmail.com +18653601377 Knoxville, TN, USA 37916 https://mmuratarat.github.io/
https://www.linkedin.com/in/mmuratarat/ https://github.com/mmuratarat

PROFESSIONAL SUMMARY

PhD-level educated, bilingual statistician and data scientist, who design, develop and implement advanced predictive methods, powered by machine learning and deep learning algorithms, in order to provide actionable insights from large volumes of real-world, structured and unstructured data in order to satisfy the business needs for decision making. I have a deep understanding of best data analytic practices to solve complex problems and I continuously improve myself and learn new emerging technologies to add new skills to my skill set. I also keep a personal blog on my webpage, where I share my technical writings/tutorials.

PROJECTS

1. **The Short-term Prediction of Electrical Load Demand of a Power System**: Applied Support Vector Regression to short-term forecast electrical load of the city Brabant of the Netherlands, comparing it with Nonlinear Auto-regressive Network and Double Seasonal Exponential Smoothing, resulting lower Mean Squared Error and the higher coefficient of determination
2. **Building/Training Long Short-term Memory Networks One Layer at a Time**: Created an LSTM cell implementation, which can be found on my Github repo, to freeze an LSTM layer and initialize next layer with the previously trained weights, or to use the pre-trained weights in a subsequent model, in order to build the network from scratch as deep as needed in real time, with decreased training time and resilience to overfitting.
3. **Sales Elasticity of Emotional Displays**: Partnered with a Direct-response TV Network, we studied over 46 million frames of video (over 12,792 hours) in which a TV host makes her/his sales pitch to customers to detect human faces and extract emotional displays using a Haar-feature based cascade classifier and mini-Xception model and matched them to customers' purchase data to understand the true nature of the relationship. We found that the sales elasticities of a host's emotional displays are uniformly negative, including that of happiness, and the presence of a host's face in the video frames is not only positively associated with sales, but also increases the sales 0.61% and, therefore, offering guidance to firms on re-training low performers on selling with a straight face as well as to reward successful hosts.
4. **Learning High-Cardinality Categorical Features with Sparse Updates**: Developed an algorithm using the key-value store idea, for high-cardinality categorical variables (e.g., ZIP-codes) without depending to the distribution of the target variable (target encoding) and without any human intervention and unrealistic expansion of feature matrix (one-hot encoding), that can exceed computational and memory requirements. It is low dimensional, thus it is very efficient in time and memory and can be computed with online learning setting in any Neural Network architecture. It also learns a weight per categorical level, which can differentiate between every unique category.

PROFESSIONAL EXPERIENCE

Graduate Teaching Assistant

Department of Business Analytics & Statistics, The University of Tennessee, Knoxville

August 2015 - Present Knoxville, TN, U.S.A.

Research Assistant

Department of Statistics, Hacettepe University

October 2011 - August 2019 Ankara, Turkey

EDUCATION

Ph.D. in Analytics (GPA: 3.92 / 4.00)

The University of Tennessee, Knoxville

2015 - 2020 Knoxville, TN, U.S.A.

NOTE: Completed my dissertation "Advances and Applications in Deep Learning" under the supervision of Dr. Michel Ballings (mb@utk.edu).

M.Sc. in Statistics (GPA: 3.93 / 4.00)

Hacettepe University

2011 - 2014 Ankara, Turkey

NOTE: Completed my thesis "A Study on Support Vector Machines" under supervision of Prof. Dr. Turhan MENTES (mentes@hacettepe.edu.tr).

B.Sc. in Statistics (GPA: 3.24 / 4.00)

Hacettepe University

2007 - 2011 Ankara, Turkey

SKILLS

Programming Languages/Libraries
- Python, Numpy, Pandas, scikit-learn, statsmodels, psycopg2, spaCy, NLTK, OpenCV and Matplotlib.

Deep Learning Libraries
- Tensorflow and Keras.

Container and Orchestration Systems
- Docker, Swarm, Kubernetes

Scripting Languages
- R, and MATLAB.

Big Data Frameworks
- Hadoop and Apache Spark

Visualization Software
- Tableau.

Statistical Softwares
- SPSS, and Minitab.

Database Systems
- SQL and relational database system, PostgreSQL.

No-SQL/Full-text Search Systems
- Elasticsearch/Kibana

Markup Languages
- LaTeX, Beamer, Jekyll, and Markdown/R-Markdown.

Version Control
- Git/Github

Miscellaneous
- Extensive usage of JupyterLab and comfortable with Linux/Unix command line and shell scripting.

Languages
- Turkish (Mother Tongue), English (Full professional proficiency), French (Elementary proficiency)