Anas Zakroum

Data Scientist

Statistics, Machine Learning, Deep Learning, Computer Perception, Decision Support Systems

+33 663 384 180

anaszkrm@protonmail.com

11 Rue des Orangers, 34090 Montpellier

Git

lvl0-statistician

Languages

- Fluent **French**
- Fluent **English** (TOEIC: 885/990)
- Native **Arabic**
- Advanced writing skills in the 3 languages

Technical Skills

Computer

- **OS**: GNU/Linux (Debian and derivatives, Red Hat Enterprise Linux, etc.).
- **Programming Languages**: R, Python, MATLAB, C, Shell scripting, HTML.
- **Web Development**: Django and Flask.
- **Database Management Systems**: SQL Server, PostgreSQL, MySQL
- **Tools**: git, docker, virtualbox, LaTeX.

Artificial Intelligence

- **Feature Engineering**: feature selection, model selection and evaluation, dimensionality reduction methods, etc.
- **Machine Learning**: Linear and logistic models, SVM, decision trees, random forest, etc.
- **Deep Learning**: dense neural networks, convolutional neural networks, recurrent neural networks, reinforcement learning, transfer learning, model fine-tuning.
- **Unsupervised Learning**: compact clustering, hierarchical clustering, autoencoders, etc.
- **Tools**: TensorFlow, Keras, Scikit-learn.

Statistics

Statistical Inference, Statistical modeling, Bayesian analysis, Exploratory data analysis, Proportional hazards, Survival Analysis, Stochastic Control, Time series Analysis, Latent variables modeling, etc.

About Me

I am interested in how statistical tools and machine learning algorithms can be used to answer questions involving transdisciplinary knowledge within data-driven decisional frameworks.

Education

2020 Present — **Master in Statistics & Data science** — Montpellier University
Stochastic processes, Exploratory data analysis, Generalized linear models, Bayesian statistics, Latent variables modeling, Time series, Stochastic control, Statistical learning, etc.

2016 2020 — **Bachelor of Science in Mathematics** — Montpellier University
General and applied mathematics.

Professional Experience

- **Development of AI tools for ecology** (March 2022 – Aug. 2022) Report
  Leveraging Deep Learning models for the detection and classification of animals appearing in images and videos in their natural habitat as a mean to perform a statistical analysis of species abundance and build occupancy models of the identified fauna. *(Data Wrangling, Feature Engineering, Deep Learning, Transfer Learning and Neural Network Fine-tuning)*

Projects

Data Analysis

- **Extreme sea waves analysis** Report
  Univariate and multivariate extreme events analysis and modeling of significant wave heights in the Mediterranean sea. *(Extreme values theory)*
- **Like mothers, like daughters?**
  Identifying patterns of resemblance among mothers and daughters lives using longitudinal data based on retrospective surveys. *(MDS, Optimal matching analysis, CA)*
- **Cellular division of E. Coli** Video
  Modeling time to division of the E. Coli bacterium using Renewal Processes and comparing the model to data. *(Renewal Processes, KDE)*
- **Polygamy risk factors** Video
  Performing multiple correspondence analysis to extract latent variables as a mean to identify risk factors of polygamy. *(Survival analysis, MCA, Cox-regression)*

Software Development

- **Worldwide Statistics Visualizer** Git repo.
  Construction of a high-level API that simplifies the visualization of statistics about countries on a world map or in different types of plots. *(package development in Python, Django framework, Model-View-Controller architecture)*

Statistical Modeling

- **Prey-predator dynamics *(Research Project)*** Report
  Constructing a stochastic version of the functional response based on renewal processes as a mean to model predatory behavior of the Canadian lynx. *(Renewal Processes)*
- **Bicycle traffic in Montpellier city** Report
  Modeling and predicting of bicycle traffic using time series. Auto regressive modeling was used to predict traffic. *(Time Series, AutoRregressive Models)*
- **Wildlife management**
  Evaluating different preservation strategies of the Gypaetus barbatus using Markov decisional processes *(Stochastic Control, MDPs)*
- **Forest density modeling**
  Modeling the arboreal density of the forest surrounding the Congo basin with various regression techniques. *(FA, PCR, PCA, PLS, Ridge, Lasso)*