# VASILEIOS VITTIS

165 Summer Street, Amherst, MA, 01002 | vvittis.github.io | github.com/vvittis | linkedin.com/in/vasilis-vittis

## education

**UNIVERSITY OF MASSACHUSETTS AMHERST** Sept 2022 - Expected Graduation 2027

Masters/PhD in Computer Science - Manning College of Information and Computer Sciences
Working with Prof. Peter J.Haas, Prof. Alexandra Meliou, and Prof. Azza Abouzied
*UMass CICS Scholarship for Best PhD Applicant 2022*
*Relevant (600- Level) Courses: Advanced Algorithms, Advanced S/W Engineering: Analysis and Evaluation, Machine Learning*

**TECHNICAL UNIVERSITY OF CRETE** Sept 2015 - Oct 2021

ECE Undergrad/ Integrated Master's Degree - School of Electrical and Computer Engineering
Class Rank: Top 4%, GPA (Computer Science): 3.6 / 4.0
***"Streaming Machine Learning in Distributed Environments for Big Data"***

**(Thesis) Random Forest Optimizations under Data Drifts**

- Designed optimizations for real-time supervised binary classification using Random Forest classifier for 25M+ real-time streaming data (**Apache Kafka)**. Utilized probabilistic distributions to model data (-95% memory), refined a state-of-the art sampling technique (-80% memory) and introduced a dynamic accuracy monitoring strategy (-70% memory).
- Performed statistical analysis to identify data drifts. Developed a distributed ensemble system on top of **Apache Flink**, using **Java** and **Scala**. Evaluated results in real time, ensuring both robustness on data drifts and scalability on Big Data.

## work experience

**Graduate Research Assistant (Database and Machine Learning)** Sept 2022 - Now

University of Massachusetts Amherst - Data systems Research for Exploration, Analytics, and Modeling Lab
***"Pushing advanced analytical models closer to database systems"***

**(Research Project) Dynamic Query Package Maintenance**

- Introduced and developed innovative algorithms to enhance **SQL** query engines for real-time package computation. Developed code for querying and efficiently managing data within a **PostgreSQL** database. Extracted data insights and created complex data visualizations using **matplotlib**. Achieved close to optimal solutions backed by mathematical guarantees using **Gurobi**.
- Improved the runtime performance through optimized data structures on large datasets, particularly useful for applications requiring immediate responses. Utilized clustering algorithms for data sampling, ensuring quick adaptability in data drifts.

## projects

**Online Credit Card Fraud Detection** - Developed a distributed fraud detection system on top of **Apache Spark.**

- Developed code for cleaning and preparing data for streaming analysis - including imbalance classes, handling missing values, formatting and normalizing. Performed exploratory data analysis on 10M+ streaming data (**Apache Kafka)** using **pandas** and **numpy**. Developed a semi-supervised binary classification model using Decision Trees. Evaluated the results using multiple standard evaluation metrics addressing issues such as: overfitting, accuracy and F1-score.
- Deployed the ML model on parallel computational nodes using **Scala** and **Java,** addressing bottlenecks and synchronization.

**Lupus (NPSLE) Classification using ML -** Analyzed high-dimensional fMRI data using ML models.

- Developed ML pipelines, used **numpy** for data computation, analyzed data distributions, investigated feature correlations, performed selection techniques for 109 features using **Python**.
- Deployed binary classification using NB, SVM and Random Forest models. Evaluated and compared each model's performance with cross-validation, ensuring low overfitting and high recall.

**ToDo AI** - Developed with **Python**, using Generative **OpenAI's GPT-3.5-Turbo Model**

- Developed a PyCharm extension which offers AI-generated code suggestions with small descriptions for code changes.

## technical skills

**Programming Languages** Python, Java, C++, SQL, Scala, MATLAB, PHP, Javascript, Solidity, R,Tensorflow

**Technologies** MySQL, PostgreSQL, Apache Flink, Apache Spark, Hadoop, Apache Kafka, Pandas, scikit-learn

## leadership experience

**IEEE TUC Student Branch** **Chairman (Sept 2020-Jan 2021)**, organized 10+ events, initiated specialized mentorship sub-teams; personally led the web development and data science group.

*all project titles are links