europass

Ivo Tavares

Nationality: Portuguese | **Phone number:** (+351) 939893474 (Mobile) |

Email address: ivo.a.v.tavares@gmail.com | **LinkedIn:**

https://www.linkedin.com/in/ivo-tavares-76781b1b9/ | **GitHub:**

https://github.com/IavTavares | **Personal Page:** https://ialvata.github.io/

ABOUT ME

Data Scientist / ML Engineer / Python Developer + PhD in Applied Mathematics.

Skilled in building and handling models and data pipelines.

Experience in using Python for NLP and Time Series forecasting tasks, in a machine learning context.

Experience in service integration and API building.

Highly flexible, deeply driven.

WORK EXPERIENCE

31/08/2021 – CURRENT

UNIVERSITY INVITED ASSISTANT PROFESSOR ISEG - LISBON SCHOOL OF ECONOMICS & MANAGEMENT

Statistics I e II (Economics, Finance and Management)

- Statistical foundations necessary for data analysis and A/B testing, e.g. point estimators, confidence intervals, hypothesis testing, etc.
- Linear Regression Model from an econometric point of view.

08/2022 – 05/2023 United States

MACHINE LEARNING ENGINEER SHAKE

- Responsible for the development of clustering algorithm (in Python), according to company, location, and software/product.
- Responsible for refactoring code and optimising algorithm performance regarding both CPU and RAM usage. (choice of design patterns, optimising use of data structures, manual garbage collection, etc.)
- Helped develop and fine tuning of NLP models, using HuggingFace Transformers package, and also PyTorch and PyTorch Lightning.
- Developed a CI pipeline using GitHub actions, and Pytest.
- Developed a Pipedream workflow to register API requests results.
- Developed an integrated Redis service as a backup and alleviate the case of race-conditions.
- Monitored in real-time test, stage and production infrastructure, using Lens (ak8 IDE), and Grafana.
- Planned sprints/cycles, using Linear.
- Tech Stack: Elasticsearch (high and low level python client),Git, GitHub Actions, Pytest, Pipedream, Scalene code profiler, Lens, Grafana, Redis, Hugging Face Transformers, PyTorch, PyTorch Lightning, Linear.

All work was developed in a remote setting.

01/2022 – 07/2022 Portugal

DATA SCIENTIST / ML DEVELOPER (REMOTE) TECTUMAI

At TectumAI, I have participated in two projects.

Digihome App (textual information extraction from pictures):

- Developed a Flask+Gunicorn+Nginx app.

- Lead on the development of a pipeline, based on a Deep Learning Algorithm for NLP / Document Parsing.
- Coordinator of the ML team, and liason with the QA team and the Frontend/Backend.
- Developed an automated CI/CD pipeline using GitLab, YAML, Pytest, Docker, and code Coverage:
    - Unit and integration tests (development servers).
    - Load, stress and soak tests, using k6 (stage servers).
    - Automated launch to production servers.
- Infrastructure AWS(S3) + GCP (DocAI).
- Performed root tasks on our AWS EC2 servers:
    - Setting repos in the servers with ssh key and 2-factor authentication
    - Creating and maintaining Docker container.
- Tech Stack: Git, GitLab, Python, Flask, Gunicorn, Nginx, HuggingFace Transformers, Spacy, Pytest, k6, Docker (creating and managing containers), AWS S3 Buckets (Python client), AWS EC2 Instances (through Linux Shell), Google Cloud Plataform for DocumentAI (Python client and Console), JIRA, Bash scripts, and VSCode IDE.

- Property price prediction:
    - Researched feature engineering.
    - Researched model comparison and finetuning.
    - Tech Stack: Git, GitLab, Python, Scikit-Learn, LightGB, OPTUNA, MongoDB, APScheduler, Google Cloud Plataform (Python client).

08/2016 – 30/08/2021 Lisbon, Portugal
UNIVERSITY TEACHING ASSISTANT ISEG - LISBON SCHOOL OF ECONOMICS & MANAGEMENT

Mathematics I and II, and Statistics I and II.

08/2013 – 07/2015 Lisbon, Portugal
UNIVERSITY TEACHING ASSISTANT NOVA SCHOOL OF BUSINESS AND ECONOMICS

Calculus I, II. (Economics, Finance and Management)

These subjects focused on univariate, and multivariate calculus, and optimization.

08/2011 – 01/2012 Carnaxide, Portugal
BUSINESS ANALYST SONAE MC

- Gathering, scrutinizing, interpreting data and analyzing results using statistical techniques.
- Creating dashboards for managerial use, showing a comprehensive overview of data from different sources helping them organizing the data according to their needs and filtering the necessary information.

EDUCATION AND TRAINING

31/08/2015 – 30/05/2021 Lisboa, Portugal
PHD IN APPLIED MATHEMATICS IN ECONOMICS AND MANAGEMENT ISEG Lisbon School of Economics and Management

Field of study Applied Mathematics and Economics | **National classification** Summa Cum Laude |

Thesis Uncertainty Quantification in State-Space models: An Example from Macroeconomics

08/2012 – 05/2015 Lisboa, Portugal
CURRICULAR APPROVAL IN PHD OF ECONOMICS NOVA School of Business and Economics

08/2009 – 04/2011 Lisboa, Portugal
MASTERS IN ECONOMICS NOVA School of Business and Economics

08/2005 – 06/2009 Lisboa, Portugal
UNDERGRADUATE DEGREE IN APPLIED MATHEMATICS AND COMPUTATION Instituto Superior Técnico - Universidade de Lisboa

DIGITAL SKILLS

Machine Learning / Data Science

Python | Natural Language Processing | Time Series Analysis | Numpy | Pandas | Matplotlib | TensorFlow | HuggingFace Transformers | PyTorch | Scikit-Learn | C++ | Java

Databases

SQL (MySQL) | PyMongo | Elasticsearch | Redis

Web Development

FastAPI | Flask, Web Scraping | HTML & CSS

Versioning

Git | GitHub | GitLab

Cloud Technologies

AWS Boto3 | Google Cloud Storage (Python Client)

LANGUAGE SKILLS

Mother tongue(s): **PORTUGUESE**

Other language(s):

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken production	Spoken interaction	
ENGLISH	C1	C2	C1	C1	C1
FRENCH	B2	C1	B1	B1	B2
GERMAN	A1	A2	A1	A1	A1
RUSSIAN	A2	A2	A2	A2	A2
SPANISH	B2	C1	B1	B1	B1

Levels: A1 and A2: Basic user; B1 and B2: Independent user; C1 and C2: Proficient user

ADDITIONAL INFORMATION

PROJECTS

19/11/2021 – CURRENT

Time Series in Python

- Developed a series of Python Jupyter notebooks (model economic fundamentals):
    - VAR and RNN: real GDP per capita and unemployment.
- Developing a Time Series package for private use in Time Series projects:
    - Classical statistical models: SARIMAX
    - Deep Learning models: LSTMs, etc.
    - Machine Learning models: Random Forest based (LightGBM, CATBoost, etc.)
    - Outliers detection: Local Outlier Factor.
    - Feature Engineering: time explicit embeddings.
    - Target/Label Engineering: stationarity.

30/06/2021 – CURRENT

Natural Language Processing

- Developed research on the potential impact of the European Central Bank speeches on economic fundamentals, using NLP techniques (BERT, LightGBM, etc.)
- Tech stack: Scikit-Learn, Keras, TensorFlow, HuggingFace Transformer.