Arnau Puche

https://github.com/puchee99

Email : arnaupuchevila@gmail.com
Mobile : (+34)-673-821-695

EDUCATION

- **Universitat Autònoma de Barcelona** Bellaterra, Spain
  Bachelor's degree of Data Engineering (240 ECTS); Avg. Grade: 7.13 *Sep. 2018 – Present*
    - **TFG:** market analysis using AI algorithms to identify patterns and trends that optimize financial decision-making.

- **Coursera**
  Data Science *Sep. 2020 – Present*
    - **IBM AI Entreprise Workflow:** connecting business priorities to technical implementations.
    - **Neural Networks and Deep Learning:** foundational concept by Andrew Ng.
    - **Fraud Detection on Financial Transactions with ML on GCP:** BigQuery and VertexAI for banking industry.

- **DeepLearning.AI**
  Data Science *April. 2023 – Present*
    - **ChatGPT Prompt Engineering for Developers:** how to use a large language model (LLM) to quickly build new and powerful applications.

EXPERIENCE

- **CaixaBank Tech** Barcelona, Spain
  Data Scientist - ML Engineer *August 2022 - July 2023*
  Tools: Python, SQL, Bash, Elastic, Gitlab,
    - **Anomaly Detection:** Detect anomalies in the failures of the entire infrastructure through different AI models.
    - **Clustering:** categorize errors using advanced clustering techniques to obtain more information.
    - **Forecasting:** prediction of different KPI.
    - **ElasticSearch:** create and modify indexes.
    - **XAI:** express why an AI system reached a particular decision, recommendation, or prediction.
    - **ML Engineer:** design self-running software to automate predictive models.

- **UAB** Bellaterra, Spain
  Research Assistant *Sept 2021 - Feb 2022*
  Tools: Python, SQL, Bash
    - **Database:** Create a SQL database from biomedical device data.

- **UVE Solutions** Manresa, Spain
  Intern Data Scientist *Feb 2021 - July 2021*
  Tools: Python, SQL, Bash, Streamlit, Pandas, Bitbucket, Confluence, Jira
    - **Data ingestion:** Build a data ingestion tool (Streamlit App) to perform ETL on different data formats. ML techniques were used to calculate similarity between words.
    - **BI:** Analyze the distribution of the hostelry sector in the Spanish territory.

FREELANCE

- **KeepUp Talent** Barcelona, Spain
  Data Scientist - ML Engineer *April 2022 - July 2023*
  Tools: Python, SQL, Bash, PowerBI, Sharepoint, Microsoft Forms
    - **App Infrastructure:** create a Microsoft Form to consult small and medium-sized companies.
    - **Business Reports:** show most relevant insight and how to improve in a PowerBI Dashboard.

HACKATHONS

- **NUWE JOBarcelona 2022 Final (7 h)** Camp Nou, Barcelona, Spain
  Data (teams of 4) - 1st *May 2022*
  Tools: Python, Github **(Repo link)**, Bash, Pytorch, Sckiit-Learn, Streamlit, Pandas, XGBoost
  Challenge: Predict the fair market value of football players.
    - **Feature Engineering:** Through different techniques and combinations with the columns, we were able to add new variables with a higher correlation.
    - **Multivariant Regression:** We use different ML models with Grid Search to find the best parameters. We also tested with neural networks.

- **NUWE CaixaBank 2022 (14 h)** online
  Data - *May 2022*
  Tools: Python, Github **(Repo link)** , Bash, Pytorch, Sckiit-Learn, Pandas, XGBoost

  Challenge: Modeling a predictive algorithm that allows knowing the future dynamics of the market from the historical data of the value of the IBEX35 and some Tweets.

  - **Sentiment analysis:** Some new columns were added in the market data. One was made by tweet's daily sentiment score. Other columns with popularity information from Google Trends.

  - **Time Series Forecasting:** Tested different Time Series and Classification models (up - down) .

- **NUWE Talent Squad 2022 (1 week)** online
  Data - 1st *July 2022*
  Tools: Python, Github **(Repo link)** , Bash, Pytorch, Sckiit-Learn, Pandas

  Challenge: The dataset was data created by sensors from the company SpaceX. The goal was to classify the data into 5 types of weather conditions in order to manage the flight.

  - **Multivariant Classification:** Neural networks were used.

- **NUWE JOBarcelona 2022 Classification (1 week)** online
  Data - Qualified for the final *April 2022*
  Tools: Python, Github **(Repo link)** , Bash, Pytorch, Sckiit-Learn, Pandas

  Challenge: Classify insects according to the value of different sensors.

  - **Multivariant Classification:** Neural networks were used.

Projects

- **Flask App**: create a functional app with different databases in the cloud. (Github link)

Programming Skills

- **Languages**: Python, Javascript, HTML, SQL, R, Bash, C, Scala

- **Technologies**: GCP, Docker, Tableau, AWS, Airflow, MongoDB, Spark, Microsoft tools

- **Libraries**: Pandas, Numpy, Flask, PySpark, Pytorch, Tensorflow, Plotly, Scikit Learn, Streamlit, Shap, Matplotlib, Request, Pymongo, Logging

LANGUAGES

Catalan: Native   **Spanish:** Native   **English:** Professional (B2 Certificate)