ÁLVARO MONFORTE MARÍN

Data Scientist | Physicist | Machine Learning Engineer | AI Developer

monfmalvaro@gmail.com | +34 696-944-160 | Bilbao (Biscay), Basque Country. Spain

varo6678.github.io | monfortemarinalvaro | varo6678 | monfmalvaro

PROGRAMMING

Python | Bash | Cython

ML & DL

PyTorch | TensorFlow | Keras | Prophet | Bi-LSTM | LLM | XGBoost | LightGBM

WEB DEVELOPMENT

FastAPI | uvicorn | Flask | Django

CLOUD COMPUTING

Google Cloud Platform | Azure

DATABASE

BigQuery | PostgreSQL | pgAdmin4 | MS SQL Server

GEOSPATIAL INF.SYSTEMS

ArcGIS

VERSION CONTROL

GIT | HugginFace

LANGUAGES

Spanish: Native

English: Professional

SOCIAL

- Organizer of the Physics Student Congress for two years: [coefis.es]
  - COEFIS VII
  - COEFIS X

ABOUT ME

Currently working in the role of data scientist in a consulting firm. I am furthering my education by studying for a Master's in Computational Physics in my spare time. With a background that includes a Master's in Data Science and a Bachelor's degree in Physics, I have transitioned from a past role as an educator to a passionate data scientist. I am enthusiastic about machine learning, deep learning, and reinforcement learning, leveraging my skills to solve complex problems and contribute to innovative solutions in the field of data science.

EXPERIENCE

Data Scientist | Lanit Consulting S.L.

09 2023 – Actually — Bizkaia, Spain

- Working on a project for a national retail company. Machine Learning. Databases. MLOps.

Collaborator | Universidad Europea de Madrid

04 2024 – Actually — Universidad Europea de Canarias, Online

- Collaborating in the creation of content for a scientific career.

Upper High School Teacher | Joshua Christian Academy

08 2022 – 05 2023 — Jacksonville, United States of America

- Physics - Honors Track, Trigonometry - Honors Track, Algebra 2, Robotics for High School and Elementary School.

Robotics Instructor | Shiva Robotics Academy

03 2022 – 06 2022 — Jacksonville, United States of America

- Teaching Elementary School children to develop in Robotics.

EDUCATION

Advanced Physics Master's degree: Computational Physics Specialty
Universidad Nacional de Eduación a Distancia (UNED)

2024 (In progress) | 60 ECTS — Portugalete, Bizkaia, Spain

Data Science Master's degree
Universitat Oberta de Catalunya (UOC)

(Finished) | 90 ECTS — Avinguda Tibidabo, Barcelona, Spain

Physics Bachelor's degree
Universidad de La Laguna (ULL)

(Finished) | 240 ECTS — S.C. de La Laguna, S.C. de Tenerife, Spain

PROJECTS

Predict molecular properties using the Graph Neural Network (GNN) framework: Infer the Scalar coupling constant from a Kaggle competition dataset | Master's thesis | UOC

(Finished) — Telematic

Initiation to the theoretical study of materials: DFT for C, Si, Ge and Sn | Bachelor's thesis | Link to project | ULL

(Finished) — Telematic

PROJECT INFORMATION

Predicting Molecular Properties Using Graph Neural Networks and Deep Reinforcement Learning: Inferring the Scalar Coupling Constant from a Kaggle Competition Dataset

This project aimed to explore the magnetic interactions between atoms and understand molecular properties through the prediction of the scalar coupling constant (SCC), a crucial parameter in the interaction between two protons. By utilizing machine learning techniques, a model was developed to analyze these interactions.

The project began with the premise that a molecule is composed of atoms connected by chemical bonds. To represent this molecular structure, graph theory was employed, with atoms as nodes and bonds as edges. This posed the challenge of accurately representing molecular structures in graphs, particularly in cases involving multiple bonds.

To facilitate the representation of molecular structures, adjacency matrices were utilized. These matrices identify connected nodes (atoms) and the degree of connection (number of bonds). A model based on Graph Neural Networks (GNN) was then designed, encompassing steps such as identifying the graph structure, specifying the type and scale of the graph, designing the loss function, and constructing the model using computational modules.

The resulting modeling system was non-directed, heterogeneous, and static, making it suitable for studying magnetic interactions and molecular properties. This project has potential applications in fields ranging from physics and chemistry to mathematics and video games.

Methodology and Infrastructure

The project followed a robust methodological process, leveraging advanced computational resources provided by Google Cloud Platform (GCP). A Virtual Machine (VM) equipped with a powerful NVIDIA Tesla P100 GPU was deployed to handle the computational demands. This high-performance GPU was crucial for managing large input tensors created through feature engineering and processing the respective GNNs via the Message Passing Neural Networks (MPNN) method.

Feature engineering involved transforming raw data into features that better represent the underlying problem for predictive models, leading to improved accuracy. The GNN was designed following the MPNN method, a novel approach that generalizes existing convolutional neural networks for processing graph-structured data. This method allowed the model to leverage both the features of the nodes (atoms) and the structure of the graph (molecular structure) for a more accurate prediction of the SCC.

In essence, the implementation of this project involved harnessing the computational power of cloud-based VMs coupled with advanced GPUs for efficient processing and training of the neural network model. This methodological approach ensured a balance between computational resources and the project's requirements, resulting in a highly accurate model for analyzing magnetic interactions and molecular properties.

Initiation to the Theoretical Study of Materials: DFT for C, Si, Ge, and Sn

This project delved into condensed matter physics, emphasizing the application of ab initio theories and computations. These calculations served as an introductory framework for understanding how various materials can be studied. The primary materials analyzed included Carbon (C), Silicon (Si), Germanium (Ge), and Tin (Sn) in their cubic and hexagonal (lonsdaleite) diamond structures. The computational tools employed included the Vienna ab initio Simulation Package (VASP), facilitating convergence studies, Equations of States using the Birch-Murnaghan approximation, analysis of Density of States, band structures, and phonon frequencies.

Given the significant computational demands of this study, collaborative efforts between research groups were vital. These collaborations enabled access to comprehensive human resources and necessary hardware and software, ensuring a more thorough analysis.

The findings of this study align with currently published calculations and theories, reinforcing their reproducibility and consistency. The results were compared with publications from various sources, predominantly Arxiv. Additionally, this research aimed to deepen the understanding of the differences between the cubic and hexagonal forms of diamond, which could be instrumental in the discovery of new materials.

Methodology and Tools Implemented

VASP was utilized to facilitate the computation of electronic properties within the framework of Density Functional Theory (DFT). This was key for the analysis, as the package enabled the study of properties of the materials, including Carbon (C), Silicon (Si), Germanium (Ge), and Tin (Sn) in both cubic and hexagonal (lonsdaleite) diamond structures. Central to this methodology was the application of the Birch-Murnaghan equations of state, computed using data extracted from VASP output files. By parsing these files, essential data such as energy, pressure, and volume were obtained. These equations allowed for understanding the relationship between volume and energy under compression. Python was further used to enhance the analysis, particularly through data interpolation, which was crucial for managing the large amounts of data generated by VASP.