

Santiago Romero

Buenos Aires, Argentina

Phone number: +11 3146 8481 E-mail: santiagoromero1234@gmail.com

Linkedin: [www.linkedin.com/in/santiago-romero-2a019a1b3]

GitHub: [https://github.com/santiagoromero20]

Machine Learning Engineer

I'm a Data Scientist, specialized in Machine Learning. I'm currently working on multiple machine learning projects with this stack of technologies: **Python, Numpy, Pandas, Scikit-learn, Keras, PyTorch, Docker and AWS**. I'm looking for my next challenge in the machine learning industry where I can find solutions to business problems and be impactful, while contributing to the development of my career.

Work Experience

Anyone AI - Machine Learning Engineer

- Products review classification: Analyze sentiment in product reviews for a movie streaming service. Manipulated data that is not in a traditional format, pre-processed it, vectorized text data using BoW and TF-IDF. Trained a word embedding and used it as a vectorization source for the data. Trained a sentiment analysis model to detect positive and negative opinions for movies reviews.
- Image Classification for E-Commerce: predict vehicle make and model from unstructured e-commerce images. Trained on a pre-built dataset of 196 classes. Visualized and cleaned the dataset, pre-processed and augmented data, and trained a fine-grained classification model using convolutional neural networks achieving 82% accuracy in the prediction of make and model combined. Deployed in AWS instances using Docker, using an API based web-service application.
- Home Credit Risk Analysis: predict whether a person applying for a home credit will be able to repay its debt or not. Manipulated and visualized data, performed data pre-processing for a large dataset of +350,000 transactions. Trained many supervised models achieving +0.72 ROC AUC. Models used where DecisionTree, XGBoost and LightGBM.
- Salary Prediction Model: The goal was to predict salary levels based on historical data for sport players. Collected and analyzed data via an API using Python and Pandas. The original data was unbalanced. Cleaned up data, and generated additional fields, stored and the created a base dataset. Manipulated and visualized data. Performed feature engineering and standarization. Selected evaluation metrics and baseline models. Trained a linear regression model, achieving an F1 score of 76%.
- Automated product Classification: The main objective of this proejct is to correctly classify a product by its name and description. Firstly, I had to analyze, clean and organize the data for further use.. As most of the data was text, I have to normalize and vectorize it with state of the art techniques, in order to deliver this data to my models. Then, perform all the

necesary training looking for the best model possible. As this was a MultiLabel Classification problem, I had to use Problem Transformation, Adapted Algorithms and Ensemble approaches to achieved that. Once I had choosen my best model, I programmed an API with Flask, so I can test the model in production.

Skills

Tech Skills: Python, Numpy, Pandas, Keras, Scikit-learn, Pytorch, SQL, Docker, AWS.

Agile Methodologies: Scrum and Kanban

Other Tools: Github.

Languages: Fluent in English, Spanish Native.

Education

National University of Buenos Aires

2021-2026

Currently studying my Bachelor's in Data Science