



# JUAN ESTEBAN CEPEDA

MACHINE LEARNING ENGINEER

## CONTACT



@juan-e-cepeda-gestion



juancepeda.gestion@gmail.com



@Juanchobanano

## SKILLS

Programming Languages: Python

Databases: RDMS, AWS Redshift, Neo4j, Data Lakehouse

Data Tools: Spark, Databricks, SageMaker

Cloud: AWS, Lambda, S3, Fargate

ML Frameworks: Tensorflow, Sklearn, PySpark

Others: FastAPI, Streamlit, Docker, Flask, Kubernetes, Git, Prophet, QA Testing, WebScraping

## EDUCATION

### M.Sc Data Science

UNIVERSITY OF HELSINKI

2023-2025

Recipient of a 100% scholarship for Master in Data Science at the University of Helsinki.

### B.S Computer Science

UNIVERSIDAD NACIONAL DE COLOMBIA

2018-2022

Dissertation work: Review of gradient policy methods in recommendation systems with the guidance of Prof. Dorota Glowacka (University of Helsinki)

### B.A Business Administration

UNIVERSIDAD NACIONAL DE COLOMBIA

2015-2020

Award: Internacional winner of Quantopian, Inc. Contest in 2020 (Algorithmic trading project)

## PROFILE

Computer scientist and business administrator with a comprehensive background spanning 5 years, encompassing machine learning modeling, software engineering, and statistical analysis. Skilled in developing advanced models for recommendation systems, time series prediction, human trafficking identification, algorithmic trading, reinforcement learning, and data engineering pipelines for diverse use cases. Experienced in collaborating with international teams across the US, UK, and Latin America.

## WORK EXPERIENCE

### MACHINE LEARNING ENGINEER

August 2023 - Present

#### YUJU APPS

August 2022 - March 2023

- Designed and developed a recommendation engine to effectively categorized product names into marketplace categories, reducing clients' time spent on categorization and improving accuracy by 9% compared to the previous baseline.
- Deployed a versatile recommender system capable of serving both batch and near real-time requests through an API. Leveraged a stack consisting of Databricks, SparkML, Tensorflow, and the OpenAI Embeddings service to meet various client use cases.
- Designed and developed time series models to predict product stock levels across multiple marketplaces. This involved implementing ETL processes to capture sales data from different marketplaces and store it in OpenSearch and Redshift.
- Implemented an alert system utilizing time series forecasting to notify clients when and where to replenish their stock, optimizing distribution across marketplaces based on expected demand.

Tech Stack: Databricks, Apache Spark, MLFlow, OpenAI API, SageMaker, SQL, Python, Git, Prophet

### MACHINE LEARNING ENGINEER

#### FACTORED.AI (BACKED BY ANDREW NG)

March 2023 - August 2023

- Served as a Machine Learning Engineer (via Factored) in a prominent US tech company based in California, contributing significantly to the creation and implementation of machine learning models in the domain of dynamic bidding estimation for advertisements.
- Conducted extensive research and development to tackle a user response prediction problem by considering demographics characteristics, advertisement types, and income variables.
- Successfully managed highly imbalanced datasets by employing advanced techniques like oversampling, undersampling, and ensemble methods. These strategies proved instrumental in effectively addressing the challenges posed by class imbalance.

Tech Stack: Tensorflow, SageMaker, Airflow, Scikit-learn, Snowflake, Apache Spark, Git

### CO-FOUNDER

#### CHAIN BREAKER

Jan 2021 - Present

- Developed a machine learning system to identify human trafficking cases in sexual ads, reducing law enforcement time spent on discovery and expediting the investigation phase.
- Collaborated with Traffik Analysis Hub ONG and Stop The Traffik to share information with financial institutions in the UK, Canada and Ireland, enhancing clients' risk assessment capabilities.
- Created multiple ETLs for web scraping and processing new advertisements from various Adult Websites Services, ensuring regular updates of information.

Tech Stack: Databricks, Apache Spark, Streamlit, Delta Sharing, Neo4j, Flask, Angular, Docker

**Keywords:** Artificial Intelligence (AI), Data Analysis, Computer Science, Deep Learning, Neural Networks, Reinforcement Learning, Natural Language Processing (NLP), Programming, Data Visualization, Data Analytics, Predictive Analysis, Statistical Data Analysis, Pattern Recognition, Data Mining