

# Isabel María Villalba Jiménez

LinkedIn: [mabelvj](#)  
Website: [mabelvj.github.io](#)

[mabelvj@gmail.com](mailto:mabelvj@gmail.com)  
+34 635 57 05 14

Github: [mabelvj](#)  
StackOverflow: [mabel-villalba](#)

## SKILLS

- Python • R
- Scraping • Unit Testing
- Bash scripting
- Docker • Docker compose
- SQL • MongoDB
- HTML • CSS
- XPath
- Regular Expressions
- Matlab

## PYTHON

- Scrapy
- Scikit-learn • Pandas
- Numpy • statsmodels
- Prophet
- Matplotlib • Seaborn
- Scrapy • BeautifulSoup
- PySpark
- Jupyter-Notebook
- Quantopian • Backtrader
- Tensorflow

## DATA SCIENCE

- Statistical analysis
- Time Series
- Deep learning

## DATA ENGINEERING

- Lambda • s3
- Kinesis • Athena
- Glue • Quicksight

## OTHER

- Agile • Scrum
- AWS
- Microsoft Azure
- Linux
- CI/CD • Open Source
- GitHub • GitLab
- Bitbucket
- $\LaTeX$  • markdown

## LANGUAGES

- Spanish** native
- English** full-proficiency  
Cambridge Certificate  
in Advanced English -  
CAE (C1) (2012)
- French** intermediate

## EXPERIENCE

### Data Engineer - SEAT:CODE *April 2021 - Present | Hybrid, full-time*

- Predictive maintenance of welding spots using Machine Learning for the SEAT Factory in Martorell.
- Implemented an ETL on AWS using Lambda, Kinesis Firehose, Glue. Data visualization using Quicksight. TDD. CI/CD using Github Actions. Infrastructure provisioning using Cloudformation.

### Data Science Engineer - Mática Partners *December 2020 - April 2021 | Remote, full-time*

- Information retrieval, Scrapers, Docker, Docker-compose, PySpark, Azure, Jenkins, AWS.
- Built an ETL to extract business information using Docker and store it in AWS s3.

### Python developer & Backend Developer - Scrapinghub *April 2019 - December 2020 | Remote, full-time*

- Information Retrieval, Scrapers, ETL development. Scrapy, Pandas, Numpy, unit testing, Pytest, Docker, Docker-compose, Celery, MongoDB, MySQL, CI/CD, cloud deployment, HTML, CSS selectors, XPath, XHR. Agile, Scrum. Code-review.

### Data Science contractor *June 2017 - April 2019 | Remote*

- **LISTedTECH** *January 2019 - April 2019 | Remote*
  - Development of a tool for retrieving information about universities. Data cleaning for database update. Scrapy. Regular expressions.
- **Windsor AI** *November 2018 - January 2019 | Remote*
  - TV attributions and ROI report generation using R and PostgreSQL.
- **Arbuckle Capital** *September 2017 - January 2019 | Remote*
  - Algorithmic trading system in (Quantopian, Backtrader (Python)). Prediction of bull-bear periods and cluster volatility periods. Used GARCH models to predict volatility. Portfolio selection.
- **SerpicoDEV** *June 2017 - September 2017 | Remote*
  - Implemented a predictor in Python to determine prices changes in commodities from markets in the U.S.

### Mentor and Reviewer Udacity

- **Classroom Mentor** *April 2017 - April 2019 | Remote, part-time*
  - \* Provided on-demand support to the **Machine Learning Nanodegree (MLND)** and the **Data Analyst Nanodegree (DAND)** students.
  - \* 120 mentees with received **average rating of 4.7**
- **Project Reviewer** *April 2017 - April 2019 | Remote, part-time*
  - \* Helped students of the **Deep Learning Nanodegree Foundation** and the **Artificial Intelligence Nanodegree** in projects related to Neural Networks, Reinforcement Learning and Statistical Analysis.
  - \* 750+ projects reviewed with received **average rating of 4.93**

### Predoctoral researcher *October 2015 - October 2016 | Barcelona* Optical Communications Group (GCO)

#### Polytechnic University of Catalonia (UPC)

- Developed Python and Matlab scripts for the simulation of optical devices
- Simulated wavelength shifter for optical networks units with 54dB side band rejection
- Designed new devices for highly efficient networks

## EDUCATION

### MSc in Photonics *2013-2014*

🇪🇺 Polytechnic University of Catalonia (UPC)  
Institute of Photonic Sciences (ICFO), UAB, UB

### Telecommunication Engineering (BSc + MSc) *2005-2012*

🇪🇺 University of Malaga

## COURSES

### Data Analyst Nanodegree , *November 2016 - June 2017*

🇺🇸 Udacity  

Certificate: [source](#)

### Machine Learning Engineer Nanodegree , *July 2016 - November 2016*

🇺🇸 Udacity 

Certificate: [source](#)

### Machine Learning Course 2016 Stanford University, Coursera

Certificate: [source](#)

## CONTRIBUTIONS

- **Open Source** - [PRs](#)

- **Scikit-Learn** :

- Fixed error when fitting RidgeCV with integers
- Added store\_cv\_values to RidgeClassifierCV and a test
- export\_graphviz should work with a sklearn.tree.\_tree.Tree.

- **Pandas** :

- BUG: Groupby quantiles incorrect bins
- BUG: Add unordered option to pandas.cut
- BUG: pivot\_table not returning correct type when margin=True and aggfunc='mean'
- DOC: update pandas.DataFrame.boxplot docstring **pandas.DataFrame.boxplot**

- **Scrapy** :

- Itemloader errors
- Raise error when start\_url found instead of start\_urls

- **Shub** :

- Hide apikey
- Insecure exception with credential should raise warning

- **StackOverflow** - [profile](#)

- Answered questions related to python, pandas, sklearn, matplotlib, numpy, ggplot
- Reputation: Top 17% overall
- Top tags: pandas, python, python-3.x, dataframe, python-2.7, pandas-groupby, matplotlib, random-forest, numpy, scikit-learn, seaborn, ggplot, r, violin-plot, machine-learning
- Top rated answers:
  - Groupby sum and count on multiple columns in python
  - Cannot make seaborn violin plot horizontal [Python3.X]
  - Adding value to only a single column in pandas dataframe
  - Print the decision path of a specific sample in a random forest classifier
  - Search optimization with pandas multi-index

## PERSONAL PROJECTS - [SOURCE](#)

- **Stocks Dashboard in Bokeh** - [source](#) *May 2018*

- Display time series automatically using Bokeh (Python).
- Easy plot arrangement and format through a dictionary of parameters.

- **Right Whale call recognition using Convolutional Neural Networks** - [source](#)