MANUEL RUEDA

New York, NY, USA

manuel.rueda@columbia.edu

(929) 235-3220

SUMMARY

Team leader and data scientist with experience on the investment management industry. My main interests are around Natural Language Processing (NLP) and Machine Learning. Check out my projects and data blog at http://masta-g3.github.io.

EDUCATION

M.S. in Data Science at Columbia University (Dec 2016)

- **Courses:** Algorithms, Machine Learning, Bayesian Models for Machine Learning, Natural Language Processing, NLP: Computational Models of Social Meaning, Data Visualization, Statistical Inference and Modeling, Distributed Computing, GIS & Spatial Analysis.

B.Sc. in Economics at Tec de Monterrey - ITESM (Dec 2009)

- **Courses:** Econometrics, Time Series, Multivariate Analysis, Statistics, Mathematics for Engineering, Linear Algebra.

Certified Financial Risk Manager (FRM) (Jan 2014, Global Association of Risk Professionals)

- **Curriculum:** Quantitative Analysis, Financial Markets and Products, Valuation and Risk Models, Risk Management, Market Risk, Credit Risk, Operational Risk, Current Issues on Financial Markets.

PROFESSIONAL EXPERIENCE

Data Science Intern at Kora Management LP. (June 2016 – Dec 2016, NY)

Kora is a global hedge fund that leverages on data science to make long term investment decisions.

- Scrapped unstructured data from e-commerce websites, aggregated it and analyzed it on a distributed system with *Spark* (*web scraping, python*, *AWS EMR*).
- Applied feature engineering on this data to build predictive models on companies' performance (*python, Spark*).
- Performed statistical analysis and modeling on proprietary datasets (*python*).

Data Science Intern at Trinnacle Capital Management (Feb 2016 – May 2016, NY)

Trinnacle is a quantitative hedge fund that uses unconventional data sources to discover investment opportunities.

- Created and administered databases with financial and market big data (*Hadoop*, *SQL, AWS*).
- Built a model to identify short-term investment opportunities based on earnings announcements. I used historical minute-frequency data for the last 10 years, and applied backtesting to find the best model parameters, such as entry and exit points (*R, SQL*).
- Created a model that uses mobile GPS data to estimate the number of customers visiting a target business, and then predict its revenue. The first test case predicted with accuracy superior to Bloomberg's estimates *(R, Shiny)*.

Senior Associate / Market Data Team Leader at MSCI (Jan 2010 – Aug 2015, MX)

MSCI provides global asset managers with tools for risk analysis and portfolio optimization.

- Started, managed and lead the local Market Data Productions team, which grew up to 8 analysts.
- Worked with Research, Engineering and Project Management to build financial risk models and deploy them to production.
- Developed a framework to cross-validate and stress test the models before deploying them to production, generating automated reports for them (*R*, *SQL*).
- Assisted global asset managers (Credit Suisse, JPMorgan, Desjardins, Nomura) in understanding these models and solve data related queries.
- Developed interactive tools and dashboards for analysts to analyze thousands of time series efficiently *(R, SQL, SpotFire*).

TA for Columbia's edX Courses (Dec 2015 – Feb 2016, NY)

- Statistical Thinking for Data Science and Analytics.
- Machine Learning for Data Science and Analytics.

MANUEL RUEDA

New York, NY, USA

manuel.rueda@columbia.edu

(929) 235-3220

DATA PROJECTS (http://masta-g3.github.io)

Linear Content Blog: Data science blog where I post tutorials and experiments on Machine Learning and NLP.

Goldman Sachs Capstone - Quantifying Changes in Financial Reports (Dec 2016): Developed a NLP tool that identifies lexical, semantic and topic changes thousands of 10-K company reports (*web scraping, nltk, python, shiny*).

Predicting Company Performance with Amazon Customer Reviews (Dec 2016)**:** Developed a machine learning based framework to analyze millions of product reviews, extract valuable information from them and analyze their relationship with different companies' stock prices.

Summer Networks (Summer 2016)**:** Several implementations of a Recurrent Neural Network that can write short poems (*numpy, theano*).

MoMA Through Time (Feb 2015)**:** Online visualization of patterns on the museum's exhibitions; awarded best of the hackathon (*R, D3*).

GENERAL SKILLS

Programming: R & Python.

Visualization: Shiny, SpotFire, & Processing (P5.js.)

DB & Cloud Computing: SQL, AWS & Spark.

Scientific Computing: pandas, nltk, sklearn, tensorflow.

GIS Analysis: QGIS, GeoDa & CartoDB.

Web: HTML, CSS & JS.