

PROFILE

Data Scientist focused on the development of machine learning and natural language processing models. Experience building data science pipelines, from data collection to visualization of results.

WORK EXPERIENCE

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

- Collected data via web scrapping, then aggregated and analyzed it on Spark.
- Applied feature engineering on this data to develop unsupervised learning models to track companies' performance.
- Statistical analysis and inference on survey data.
- Developed Shiny apps to monitor data collection processes and analyze model data.

Data Science Intern @ Trinnacle Capital Mgmt. (Feb – May 2016, NY)

- Used non-traditional sources of data to model companies' revenue.
- Used feature engineering and cross-validation to refine the model results.
- Developed interactive Shiny dashboards to explore and visualize the results.

Senior Associate (Market Data Productions) @ MSCl (Jan 2010 – Aug 2015, MX)

- Built and managed the Market Data Team, leading 3 analysts.
- Developed interactive Spotfire dashboards used by analysts to perform daily QA on >10,000 time series.
- Developed automated reporting scripts in R for risk models validation.
- Coordinated projects with the global Research and Project Management teams.

DATA SCIENCE PROJECTS (masta-g3.github.io)

Geospatial Analysis of Gentrification in NYC (Dec 2015): Geospatial analysis of the gentrification phenomenon on the NYC area (*R*, *QGIS*, *CartoDB*).

CULPA Sentiment (Feb 2015): Sentiment analysis and visualization of academic course ratings. (*Python*, *HTML*, *D3*).

MoMA Through Time (Feb 2015): Visualization of exhibitions patterns on the museum; awarded best of the contest (*R*, *HTML*, *D3*).

Summer Networks (Summer 2016): Several tutorials and experiments on Neural Networks, on my attempt to build a RNN that can write short poems.

ACADEMICS

M.S. in Data Science (Dec 2016, Columbia University, GPA 3.4).

- Relevant Courses: Algorithms, NLP, Machine Learning, Bayesian ML, Data Visualization, GIS Spatial Analysis, Distributed Computing.
- Capstone Project (Goldman-Sachs): Use NLP models to identify and quantify changes on SEC reports.

B.Sc. in Economics (Dec 2009, Monterrey Tech, GPA 3.5).

- Relevant Courses: Econometrics & Time Series, Multivariate Analysis, Statistics, Mathematics for Engineering, Linear Algebra.
- Published Social Evaluation of the Caritas' Food Bank Program at ITESM's Journal of Economics (Year 2, Vol. 2, 2009).

CONTINUING EDUCATION

Global Association of Risk Professionals (Jan 2014).

- Certified Financial Risk Manager (FRM).

John Hopkins University (March – Sept 2014).

- Online Courses: Data Scientist Toolbox, R Programming, Exploratory Data Analysis, Getting and Cleaning Data.

VOLUNTEER WORK

TA for Columbia's edX courses:

- Statistical Thinking for Data Science and Analytics (Dec 2015 – Jan 2016).
- Machine Learning for Data Science and Analytics (Jan – Feb 2016).

KEY SKILLS & COMPETENCIES

Visualization: SpotFire, Shiny, D3 & P5.js.

Programming: R & Python.

DB & Cloud Computing: SQL, AWS & Spark.

GIS Analysis: QGIS, GeoDa & CartoDB.

Web: HTML, CSS & JS.

MAIN INTERESTS

Machine learning, natural language processing, web scraping, data apps & dashboards.