

CESAR ACOSTA

cesar-acosta.github.io | linkedin.com/in/cesar-acosta | cesar.acostamejia@gmail.com

More than 10 years of experience analyzing highly complex data, building advanced machine learning models to predict market outcomes useful to improve decision making in Marketing Analytics, Financial investing, and business operations analytics. Outstanding skills to finding insights for problem solving and process optimization.

EDUCATION

- Ph.D. Statistics, The University of Texas at Dallas
- MBA, ITESM Campus Ciudad de México

SKILLS

- Predictive Analytics: Machine Learning - Python
- Deep Learning with Python
- Prescriptive Analytics with Simulation
- Advanced Statistical Analysis
- Data Mining with Python, R
- Data Visualization: *R* ggmap, Tableau

Ample experience applying Machine Learning to solve real-world problems through the use of advanced statistical analysis to build prediction and classification models in applications such as Marketing Research, Portfolio Optimization, Volatility Forecasting, and Automated trading. Experience using Deep learning, Ensemble Methods, Gradient boosting, and other classification methods (KNN, Discriminant Analysis), Regression methods (Ridge, LASSO, Logistic, Multinomial regression), and Clustering methods (PCA, K-means, Hierarchical clustering).

PROJECTS

- A marketing strategy required to group a large set of US metropolitan cities into clusters. It was deemed critical to characterize the cities in each of the clusters. A K-Means algorithm was used to cluster the cities and a biplot and pivot tables were used to characterize the cities in each cluster. The project was completed with python. It was found which attributes the cities in each cluster have in common and how much the cities from different clusters are distinct. This solution helped the Director to efficiently allocate resources and marketing efforts to the different clusters attaining the best possible revenue per dollar invested. March 2024.
- A health insurance company asked for help to estimate the medical expenses charged to their top insurance plan for the next year. The manager wanted to predict how much more expensive is the premium for each additional children and how much the medical expenses increase for each additional year of age. To this end a nonlinear regression model is developed using the patient's age, sex, body mass index (bmi), number of children, smoker (yes or no), and location as predictors. Some useful interactions are found that improve the model accuracy. The results allowed the company answer many business questions such as how much more expensive is a smoker than a non-smoker, and how much more expensive are obese smokers than non-smoker patients that are not obese. August 2023.
- The client customer data from a large number of emails is used to predict if they are spam. Classification models were built to find the most accurate predictions. 57 variables indicating the frequency of certain words in the email were used in the models. It is expected that these words are useful for prediction. 5-fold cross validation was used to estimate and compare the models performances. The best model attained a 94% test accuracy rate. May 2022.
- Mixtures-based Value at Risk Estimates of Financial Stocks. Value at risk (VaR) and Conditional VaR (CVaR) are two common measures of risk that are related to the loss distribution. It is generally believed that if the true loss distribution is heavy-tailed, as compared to the normal, then the risk is higher. It was shown that in general this is not the case. Formulas for VaR and CVaR for mixtures were derived and it was shown that there are instances where the normality assumption overestimates (and the mixture distribution underestimates) the observed market risk. January 2020.
- Improved an S&P 500 Investment tracking portfolio by changing the assets allocation to reduce the associated risk. Investment portfolios with and with no assets' shortselling were derived. An optimization model with cardinality constraints was considered for each case. The portfolio's risk was reduced in term of the return's volatility. Portfolios based on other risk measures such as inverse volatility portfolio, equal-risk-contribution portfolio, and, maximum diversification portfolio, were considered. December 2019.
- Compared the performance of Random Forest with Multinomial Regression to classify customers into several market segments. It was shown that Random Forest and Ensemble methods outperform multinomial regression models in terms of predictive performance. July 2018.
- Optimized Marketing Resource allocation to improve customer acquisition and retention. The nonlinear relation between marketing efforts and customer acquisition and retention was derived. Values of annual spending on acquisition and retention that maximize return on investment were found. July 2017.

EXPERIENCE

- Data Scientist Consultant 2013 - 2024
- USC, Viterbi School of Engineering, MS Analytics, program director 2018 - 2024
- USC, Viterbi School of Engineering, MS Analytics, Predictive & Prescriptive Analytics, instructor 2016 - 2024
- USC, Viterbi School of Engineering, MS Analytics, Data Mining, instructor 2016 - 2024

INVITED SPEAKER

- IDEAS 2019 Conference on AI. *Data Science and Analytics. Competing in a data-driven World*. International Data Engineering and Science Association. October 2019, Los Angeles, CA.
- SatRday LA 2019 *Multiple Response Regression Models*. Los Angeles R Users Group. April 2019, Los Angeles, CA. <https://losangeles2019.satrdays.org/>
- IDEAS 2018 Conference on AI. *Is the Best Predictor actually the best?*. International Data Engineering and Science Association. October 2018, Los Angeles, CA. www.ideassn.org/socal-2018/

PUBLICATIONS

- Acosta-Mejia, C. A., Rincon, L. A., "The Continuous Run Sum chart", *Communications in Statistics - Theory and Methods*, 43: 4371 - 4383, 2014.
- Acosta-Mejia, C. A., "Two-sided charts for monitoring nonconforming parts per million", *Quality Engineering*, 25, pp. 34 - 45, 2012.
- Acosta-Mejia, C. A., "On the Performance of the Conditional Decision Procedure in Geometric charts", *Computers and Industrial Engineering*, 61, pp. 905 - 910, 2011.
- Acosta-Mejia, C. A., Pignatiello J. J., "The Run Sum R chart with fast initial response", *Communications in Statistics - Simulation and Computation*, 39, pp. 921 - 932, 2010.
- Acosta-Mejia, C. A., Pignatiello J. J., "ARL-Design of S Charts with k -of- k Runs Rules", *Communications in Statistics - Simulation and Computation*, 38, pp. 1625 - 1639, 2009.
- Acosta-Mejia, C. A., Pignatiello J. J., "Modified R charts for improved performance", *Quality Engineering*, 20, pp. 361 - 369, 2008.
- Acosta-Mejia, C. A., "Two sets of runs rules for the \bar{X} chart", *Quality Engineering*, 19, pp. 129 -136, 2007.
- Acosta-Mejia, C. A. Pignatiello J. J., "Monitoring the Variability of Symmetric Processes", *International Journal of Industrial Engineering*, 9, pp. 151-161, 2002.
- Acosta-Mejia, C. A., Pignatiello J. J., "Monitoring Process Dispersion with no Sub-grouping", *Journal of Quality Technology*, 32, pp. 89-102, 2000
- Acosta-Mejia, C. A., Pignatiello, J. J., Rao, V. B., "A Comparison of Control Charting Procedures for Monitoring Process Dispersion", *IIE Transactions*, 31, pp. 569-579, 1999
- Acosta-Mejia, C. A., "Improved p charts to Monitor Process Quality", *IIE Transactions*, 31, pp. 509-516, 1999
- Acosta-Mejia, C. A., "Monitoring Reduction in Variability using the Range", *IIE Transactions*, 30, pp. 515-523, 1999

AUTHORED BOOK

Financial Derivatives, 2018. My textbook for a course in financial derivatives, portfolio optimization, and hedging. The book includes examples and exercises in R to construct optimal portfolios, to estimate Value at Risk, to price European and American options, among other applications. It also introduces Stochastic processes and stochastic calculus for the Black and Scholes formulas, and covers Monte Carlo simulation of Brownian motion to estimate the price of some exotic options. It shows how to use libraries `RQuantlib`, `Rmetrics`, `rugarch`, `fOptions`, `fExoticOptions` for financial modeling.

AWARDS

- 2018-2019 Outstanding Teacher of the Year, USC Department of Industrial and Systems Engineering
- 2018, Best FE Track Paper Award. Castro R., Huang S., Liu J., Blay R., Acosta-Mejía C. *Mixtures-based Value at Risk Estimates of Financial Stocks*. Third North American International Conference on Industrial Engineering and Operations Management, IEOM Society International.