Eric Cruz

New York (Open to Remote) | (914) 589-4022 | eric.cruz.4242@gmail.com | linkedin.com/in/eric-cruz-07124651 | github.com/cruzer42

EXPERIENCE

E CRUZ LLC **New York, NY**
Data Analyst *December 2019 - present*

- Working on company-sponsored Data Science projects in the Apziva residency program, having completed the Springboard Data Science Career Track certification while continuing to work full time.
- As a Regulatory Technology Scrum Master in Agile Transformation at Wells Fargo, coached a new team to achieve target metrics using data-driven results to optimize sprint planning, resource allocation, and delivery of business value to stakeholders.
- After a successful first engagement, achieved next level targets while simultaneously coaching 3 new teams of approximately 7 developers per team.
- Overall Predictability measures in the optimal range of 85% to 115%, over 91% of the time. Agile Maturity Level 2 achieved, with significant progress being made toward Level 3.

Dexia Credit Local **New York, NY**
Project Manager in IT and Data Quality Services *July 2014 – December 2019*

- Regulatory Projects covered CFTC, EMIR, MIFIR SFTR for Trade Repositories DTCC and UnaVista. In addition to PM role, I covered the Business Analyst role and reconciled trade details and population matching, comparing SQL database queries to Excel lookup, pivot, and sorting functions.
- System Upgrade Projects covered Murex, Principia, and FIS Prime Compliance Suite. All were delivered on time and within budget.
- Analyzed and rewrote specifications for a Bond Pricing GUI based tool, sourcing data from 3 external sources to comply with best practices. Managed project to successful delivery of the product.

HSBC **New York, NY**
IT Consultant – o3 Technology Solutions *February 2008 – June 2014*

- Contributed and consulted as IT Business Analyst, delivering support derivatives software Misys Summit. Supervised activities of 3 to 6 Developers and 3 to 6 Analysts on projects and ran/maintained/changed the bank activities. Documented Business Requirements, Project Plans/Timelines, SIT/UAT Test Plans, Specifications, Status Reports.
- Projects and support were delivered within communicated timeframes and budget 100% of the time, which led to multiple contract extensions and additional project hiring.

PROJECTS

Alumni Network Social Outreach Tool
https://github.com/cruzer42/Capstone-Project-2-for-Springboard

- The larger goal is to find specific people on social media and contact them on behalf of an alumni organization to which they belong. The data collection is intended to be personally reviewed and supplemented, because additional data for sources like Facebook, Instagram, Twitter etc., have technical barriers preventing searches for specific people.
- For a given college tennis website, scrape the name and roster info for all available years, and supplement with other documents such as all-time or historical records in downloadable format. After compiling the full roster, search for each name on LinkedIn, including school name and "tennis" in the query. Download the profile and predict whether the correct person is found. This automated search is performed using a Selenium algorithm to simulate the pace that a fast person could achieve, to avoid bot detection.
- The baseline data was collected as part of a specific team's general outreach campaign, and test data was collected for several other rosters. The nature of the website makes it conducive to alumni-to-alumni outreach,

so the real value of the tool is to encourage completion of the chain by identifying cohort leaders who will supplement the LinkedIn data with information from other sources.
- The data collected is meant to be shared amongst the various teams, to connect the community from the ground up with natural hierarchies such as conferences, regions, divisions that can be used to expand participation and renew rivalries in a social setting. The roster list is valuable for starting the outreach, as roughly half the people across age groups are found successfully on LinkedIn. The data is intended to be used as feedback to re-train and improve the model.

Predicting Customer Satisfaction
- Built a random forest model to garner insights about a customer satisfaction survey, regarding predictive power of each feature, and the minimum feature set required to surpass the 73% accuracy target.
- Adjusted the survey response scale to be centered around zero to capture the nature of positive/negative sentiment. Introduced the sum/average as an additional feature for a consolidated measure of overall satisfaction. This improved the explanation and result compared to reducing features on the dataset of 6 questions and 126 Responses.
- Utilized automated model selection and parameter selection routines, optimized the train, test, split ratio, and considered additional measures of fit appropriate for the size of the dataset. Achieved a recall score above 90% in the final model, as a measure of focusing on the ability to identify all the positive responses.

Term Deposit Marketing
- Given a dataset of 14 features, the client would like to predict if a customer will subscribe to a term deposit marketing campaign, in order to segment the customer population by likelihood to subscribe. The features are a mix of binary, categorical, and numerical. The response is highly imbalanced, with a success rate for subscriptions sold of about 8%.
- Performed extensive visual analysis of the features after converting categorical features to binary, to examine interesting correlations and feature importance relationships in the expanded view. Selected PCA for pre-processing pipeline to address dimensionality reduction after expanding categorical features and SMOTE to address imbalance. This combination improved the model fit while maintaining the accuracy target of 81%.
- Recommended a significant reduction in the call list, along with evidence to support a data driven Decision Support System (DSS) for targeted marketing. The time savings are significant, as inverse of the population success rate implies that 92% of the call list can be ignored. The recommendation includes providing feedback, or using actual results to continuously train the model.

Word Search Similarity
- Given a list of individuals with job title and other info, the client would like to use keyword search comparisons to identify and rank candidates based on job title. After a manual review of a candidate is conducted, re-ranking is conducted with a starring and recalculating routine.
- Cleaned and vectorized the job title data using normalization, removal of stop words and special characters, tokenization, and lemmatization. Compared similarity scores using several publicly available pre-trained word vectors, representing a variety of word, phrase, and context embedding techniques.
- Analyzed the impact of re-ranking and keyword variation on this Natural Language Processing (NLP) challenge. Reduced the candidate list from 51 to 7, an 86% reduction representing time saved.

Image Classification
- Computer Vision (Reviewing labelled dataset and re-training with updated labels)
- Did a comprehensive analysis of the dataset and manually corrected mislabeled data.
- Worked with CNNs to predict whether a page flipping action is present in an image taken from a video sequence.
- Trained a Convolutional Neural Network using PyTorch framework. Improved the baseline performance by 10%.
- Did a post-training analysis to understand which of the examples the model is failing to predict to improve the model performance.

EDUCATION

The University of Chicago – Booth School of Business **Chicago, IL**
MBA in Finance and Econometrics *1995*

Cornell University **Ithaca, NY**
AB in Economics and Philosophy *1987*

LICENSES & CERTIFICATIONS

- Springboard Data Science Career Track: 550+ hours of hands-on curriculum, with 1:1 industry expert mentor oversight, and completion of 2 in-depth capstone projects. Mastering skills in Python, SQL, data analysis, data visualization, hypothesis testing, statistics, and machine learning.
- Coursera: Deep Learning and Big Data Specialization series
- Project Management Professional (PMP), Certified ScrumMaster (CSM), DTCC GTR Certification

SKILLS & INTERESTS

Skills: Python, SQL, Excel, data analysis, data visualization, hypothesis testing, machine learning, AI, deep learning, NLP, BERT, word2vec, GloVe, FastText, PyTorch, Scikit-learn, optimization, Tableau, JIRA, regression analysis

Interests: Tennis, Platform Tennis, Skiing, Hiking