


Cesar Contreras	
Machine Learning Engineer/Data Scientist	
P: 404-543-6716	
G: xxxxx@gmail.com	

Summary

- 7 years experience in Data Science and Business Management Consulting utilizing Machine Learning, Deep Learning, Statistical Modeling, Data Mining, Data Visualization, and Data/Business Analytics.
- Transform business concepts and needs into mathematical models, design algorithms, and build and deploy custom business intelligence software solutions.
- Work full software development lifecycles and meet benchmarks within given deadlines with Agile and Scrum methodologies.
- Build models with deep learning frameworks such as TensorFlow, PyTorch, and Keras.
- Well-versed in advanced statistical and predictive modeling techniques to build, maintain, and improve on real-time decision systems.
- Quickly adapt and understand new subject matter domains and design and implement effective novel solutions to be used by other subject matter experts.
- Experience in applying statistical analysis and machine learning techniques to live data streams from big data sources using Spark and Scala.
- Proficient in the application of statistical learning methods including regression analysis, forecasting, decision tress, random forest, classification, cluster analysis, support vector machines, and naive Bayes techniques.
- Analyze large datasets on distributed databases and develop machine-learning algorithms to gain operational insights and present them to leadership.
- Proficient using statistical tools and programming languages such as Python, R, **LINUX/UNIX**, SQL (multiple flavors), MATLAB.
- Clearly present complex concepts to technical or non-technical audience by incorporating data visualizations and widgets created in Python, R and Tableau (Matplotlib, Seaborn, Ggplot2, and Plotly).
- Experience with handling and creating statistical models mostly with python and TensorFlow on big data sets using cloud/cluster computing (AWS cloud computing, Google Cloud Platform, Azure, Spark).
- Experience using SQL queries to work with and manipulate relational databases.

Comentado [A1]: I don't think I will performce well in these

Technical Skills Table

<ul style="list-style-type: none"> Libraries - NumPy, SciPy, Pandas, Theano, Caffé, SciKit-learn, Matplotlib, Seaborn, Plotly, TensorFlow, Keras, NLTK, PyTorch, Gensim, Urllib, BeautifulSoup4, PySpark, PyMySQL, SQLAlchemy, MongoDB, sqlite3, Flask, Deeplearning4j, EJML, dplyr, ggplot2, reshape2, tidyr, purrr, readr, Apache, Spark. 	<ul style="list-style-type: none"> Programming Languages - Python, R, SQL, Java, MATLAB, Mathematica, C#, C++, Javascript Applications - Machine Language Comprehension, Sentiment Analysis, Predictive Maintenance, Demand Forecasting, Fraud Detection, Client Segmentation, Marketing Analysis
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Comentado [A3]: I had peculiar problems learning C++



<ul style="list-style-type: none">Machine Learning - Supervised Machine Learning Algorithms (Linear Regression, Logistic Regression, Support Vector Machines, Decision Trees and Random Forests, Naïve Bayes Classifiers, K Nearest Neighbors), Unsupervised Machine Learning Algorithms (K Means Clustering, Gaussian Mixtures, Hidden Markov Models, Auto Encoders), Imbalanced Learning (SMOTE, AdaSyn, NearMiss), Deep Learning Artificial Neural Networks, Machine PerceptionAnalytics - Data Analysis, Data Mining, Data Visualization, Statistical Analysis, Multivariate Analysis, Stochastic Optimization, Linear Regression, ANOVA, Hypothesis Testing, Forecasting, ARIMA, Sentiment Analysis, Predictive Analysis, Pattern Recognition, Classification, Behavioral ModelingNatural Language Processing Document Tokenization, Token Embedding, Word Models, Word2Vec, FastText, Bag Of Words, TF/IDF, Bert, Elmo, LDA	<ul style="list-style-type: none">Cloud - Analytics in cloud-based platforms (AWS, MS Azure, Google Cloud Platform)Quality - Continuous improvement in project processes, workflows, automation and ongoing learning and achievement CLOUD Analytics in cloud-based platforms (AWS, MS Azure, Google Cloud)Development - Git, GitHub, GitLab, Bitbucket, SVN, Mercurial, Trello, PyCharm, IntelliJ, Visual Studio, Sublime, JIRA, TFS, Linux, UnixData Extraction and Manipulation Hadoop HDFS, Google Cloud Platform, MS Azure Cloud, SQL, NoSQL, Data Warehouse, Data Lake, SWL, HiveQL, AWS (RedShift, Kinesis, EMR, EC2, Lambda)
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Comentado [A2]: Coud it be Bagging (I already work with that in Rstudio)

Project Work Experience

Computer Vision Specialist

February 2020 to Current

Micron, Austin

Micron is a well-integrated semiconductor-based company with foundries and fabs all over the World. Micron specialize in planar based multi-stack systems. Their Austin facility hosts automotive, memory and institutional research. Their principal problem is the detection and forecasting of Angstrom-scale device failure. In order to solve this problem several solutions were implemented. First a combination of logistic regression and decision trees were used to classify failure based on various parameters during the production process. Finally, a machine vision stage was set up to detect physical visible error, using a convolutional neural network to verify production stages and aid in feature engineering for regression stages.

- Developed a predictive model and validate Neural Network Classification model to facilitate prediction algorithms.
- Improved efficiency of the model by boosting method on prediction model to improve efficiency.
- Used Convolutional Neural Networks and Machine Vision to detect and predict flaws in stereo lithography nano machined stacks.
- Used R and Python for programming for improvement of model and explored regression and ensemble models in machine learning to perform prediction.
- Developed a predictive model and validated Neural Network Classification model to predict the feature label.



- Developed machine learning algorithms utilizing Caffe, TensorFlow, Scala, Spark, MLlib, R SciPy, Matplotlib, NLTK, Python, SciKit-Learn, etc.
- Performed statistical analysis and built statistical models in R and Python using various supervised and unsupervised Machine Learning algorithms like Regression, Decision Trees, Random Forests, Support Vector Machines, K- Means Clustering and dimensionality reduction.
- Used MLlib, Spark's Machine learning library to build and evaluate different models.
- Transformed logical data model to physical using ERwin ensuring the primary key - foreign key relationships, consistency of definitions of data attributes and indexes.
- Designed the Data Marts in dimensional data modeling using star and snowflake schemas.
- Used Erwin for effective model management of sharing, dividing and reusing model information and design for productivity improvement.
- Worked with project team representatives to ensure that logical and physical ER/Studio data models were developed in line with corporate standards and guidelines.
- Define the list codes and code conversions between the source systems and the data mart enterprise metadata library with any changes or updates.
- Developing and enhancing statistical models by leveraging best-in-class modeling techniques.
- Developed a predictive model and validated Neural Network Classification model for predicting the feature label.

Senior Data Scientist

September 2018 to February 2020

Index Fresh, Bloomington CA

Index Fresh is worldwide marketer of avocados, sourcing from all major growing regions around the globe. To ensure appropriate procurement of avocados, Index Fresh needed to build a image recognition tool to help their growers identify potential diseases on avocado trees. I was tasked to build a classification model using a Keras Convolutional Neural Network that can take in photos from field workers, to then returns specific disease classification. The model utilized transfer learning from Densenet201v2, and my demo had a accuracy rate of 93%. After finishing the model training, the weight were stored, and shared with the deployment team.

- Used SQLAlchemy to perform queries and pull data from Amazon S3 MemSQL database into Pandas DataFrames in Python.
- Created python scripts that converts images to a jpeg filetype then resizes to a set size
- Utilized Tensorflow's Keras Image Processing to develop ImageDataGenerator objects to augment(rotate, blur, flip) images for a Convolutional Neural Network
- Imported from Tensorflow Kera Applications InceptionResNetV2 as the base model for the initial CNN using sample images from field workers and online resources
- Utilized a second pretrained CNN called DenseNet201 as transfer learning technique to train again data collected from the MemSQL database
- Developed basic Flask app using a static HTML template as a UI to upload images, then return a processed version image with a disease classification

-
- Evaluated training data, and testing data using SOUS, along with various over sampling, and under sampling techniques to placate unbalanced image data.
 - Coordinated with field workers, logistic warehouse workers, and other departments to ensure the image recognition product fits the Avocado disease use case in deployment
 - Defined different metrics and indicators for ensure the CNN model maintains its integrity

Data Scientist

March 2017 to September 2018

Santander Bank, Boston(Remote)

Santander Bank is a diversified financial business that helps people and businesses prosper. I was part of a small team of data scientists that worked with the Security Department tasked with fraud detection. I applied an ensemble of classification and unsupervised models to identify fraudulent activity within the incoming transactions. Once deployed, we were able to reduce the number of incorrectly identified fraudulent charges and save our customers time and money.

- Stratified imbalanced data to ensure fair representation of the minority data in all data sets used for cross validation of the model.
- Consulted with regulatory and subject matter experts to gain clear understanding of information and variables within data streams.
- Utilized cloud computing resources for model optimization/tuning of hyperparameters, and cross-validation of statistical data science models.
- Extracted data from Hive database on Hadoop using Spark through PySpark.
- Used R's dplyr for data manipulation, as well as ggplot2 for data visualization and EDA.
- Utilized Scikit-Learn, SciPy, Matplotlib, and Plotly for EDA and data visualization.
- Built Artificial Neural Network models to detect anomalies using PyTorch and Scikit-Learn.
- Used Scikit-Learn's model selection framework to perform hyper-parameter tuning using GridSearchCV and RandomizedSearchCV algorithms.
- Developed unsupervised K-Means and Gaussian Mixture Models (GMM) from scratch in NumPy to detect anomalies.
- Employed a heterogeneous stacked ensemble of methods for the final decision on what transaction was fraudulent.
- Deployed model using a Flask app stored in a Docker container.
- Evaluated the performance of our model using a confusion matrix, accuracy, recall, precision, and F1 score. Took careful consideration of the recall score.
- Utilized Git for version control on GitHub to collaborate work with the team members.

Junior DataScientist

July 2015 to March 2017

Jifkins, Mexico City

JIFKINS is an agrochemistry company who prepare waxes for vegetables, fruits and make some analysis for the farm field. I was recruited to detect a wasting material problem with the elaboration of one wax that is applied to pineapples. So, for the analysis a multilinear regression model was used, after having a 88% score, where the residuals showed it had 2 behaviors based on the temperature.

- Worked in Git development environment.
- Applied expert-level Python and SQL Server development skills.
- Utilized Tensoflow, Keras, Python, and deep neural network and analytical techniques.
- Transformed business requirements into analytical models, designed algorithms, built models, and developed data mining and reporting solutions that scaled across massive volumes of structured and unstructured data.
- Applied Statistical NLP / Machine Learning, especially Supervised Learning- Document classification, information extraction, and named entity recognition in-context.
- Worked with Proof of Concepts (POC's) and gap analysis and gathered necessary data for analysis from different sources.
- Prepared data for data exploration using data wrangling.
- Designed physical data architecture of new system engines.
- Implemented neural networks.
- Worked with Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, neural networks, Principal Component Analysis, and Recommender Systems.
- Developed Logical and Physical Data models and organized data per the business requirements using Sybase Power Designer, ER Studio in both Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) applications.
- Designed Star Schema and Snow Flake schema for Data Warehouse and Operational Data Store (ODS) architecture.
- Worked with Python and Scala languages and software packages such as Stata, SAS, and SPSS to develop neural network and cluster analysis.
- Designed visualizations using Tableau software and published and presented dashboards,

Education

Bachelor's (Nanotechnology Engineering) - ITESO

Certifications

AWS Architecture , Core Services.