

TeraCrunch Socratez™ Platform & Methodology

DATA PREPARATION

Identifying fields for the model, conversion of categorical data types to numeric types, featurization of text data, and joining of relevant tables in a relational database. Data imputation for missing values

FEATURE ENGINEERING

Feature engineering is the process of using knowledge of the data to create features that make machine learning algorithms work. Coming up with features is difficult, time-consuming, requires expert knowledge.

MODEL TRAINING

Explores & identifies best options from a range of machine learning models (generalized linear models, decision trees, random forests, gradient boosted decision trees and neural networks).

TEST & SIMULATION

Explores changes in predictions if inputs are changed. This allows exploration of the underlying causal effects in the model.

For more information contact:
advancedanalytics@teracrunch.com

Case Study

Construction of a descriptive model for the main barriers of entrepreneurship within the United States

About The Client

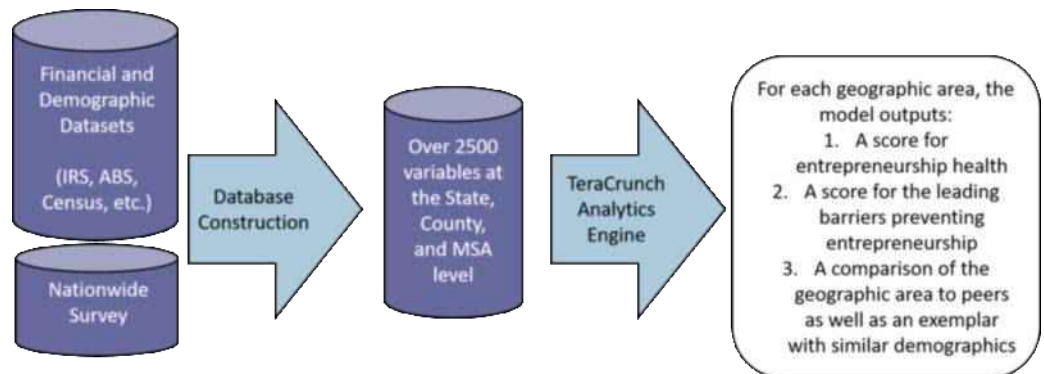
A leading non-profit in the area of entrepreneurship

Problem Overview

Entrepreneurship is a mainstay of the American economy, as well as a symbol of the "American Dream". Despite this, researchers struggle to properly evaluate the complex cultural, financial, and legislative practices which foster healthy entrepreneurship, and the barriers which stifle it.

TeraCrunch Solution

- Drawing from numerous high quality public datasets, we have constructed a dataset of financial and demographic data which can be broken down at numerous levels (State, County, MSA, etc.)
- A nationwide survey of entrepreneurs, potential entrepreneurs, and failed entrepreneurs was constructed for the client to provide reliable insight into the primary barriers to entrepreneurship
- From the large, reliable dataset which has been constructed, we have produced a smaller dataset of the most important variables to entrepreneurship
- Machine learning models were tested and deployed using best practices to produce the desired products
- As this is an area of fundamental research within the area of entrepreneurship, we have provided guidance in the area of data science and robust research practices within the area of data handling which has molded the final product and the path to its construction



Impact on the business

Our solution has produced a high quality, large database for use by researchers within their organization. This data will prove monumental in changing the direction of research for years to come. Additionally, we are producing robust, defensible reports from our model which are providing publicity and interest to researchers, entrepreneurs, and lawmakers across the country.

The output of the model allows us to produce an "Index of Barriers to Entrepreneurship" which will provide a basis for stakeholders to make decisions. Additionally, it allows us to better produce future models which will give more predictive and prescriptive information regarding entrepreneurship across the country.