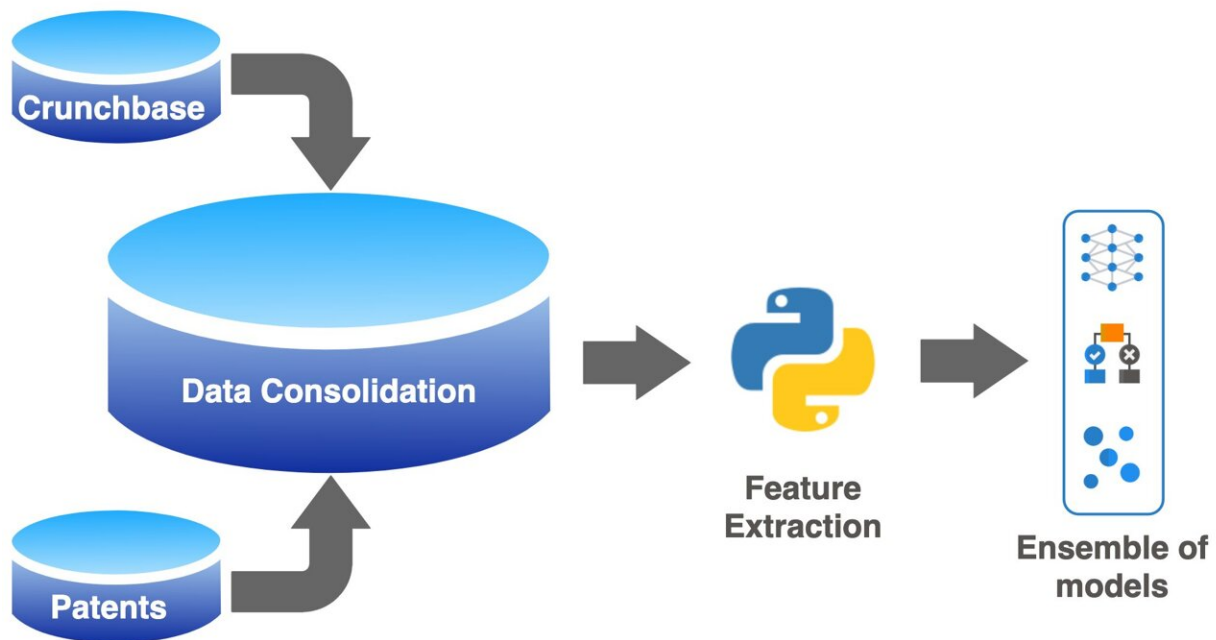# Scientists develop AI to predict the success of startup companies

September 7 2021

The machine-learning pipeline used to train the models. Credit: Greg Ross

A study in which machine-learning models were trained to assess over 1 million companies has shown that artificial intelligence (AI) can accurately determine whether a startup firm will fail or become successful. The outcome is a tool, Venhound, that has the potential to help investors identify the next unicorn.

It is well known that around 90% of startups are unsuccessful: Between 10% and 22% fail within their first year, and this presents a significant risk to venture capitalists and other investors in early-stage companies. In a bid to identify which companies are more likely to succeed, researchers have developed [machine-learning models](#) trained on the historical performance of over 1 million companies. Their results, published in KeAi's *The Journal of Finance and Data Science*, show that these models can predict the outcome of a [company](#) with up to 90% accuracy. This means that potentially 9 out of 10 companies are correctly assessed.

"This research shows how ensembles of non-linear machine-learning models applied to [big data](#) have huge potential to map large feature sets to business outcomes, something that is unachievable with traditional linear regression models," explains co-author Sanjiv Das, Professor of Finance and Data Science at Santa Clara University's Leavey School of Business in the US.

The authors developed a novel ensemble of models in which the combined contribution of the models outweighs the predictive potential of each one alone. Each [model](#) classifies a company, placing it in one of several success categories or a failure category with a specific probability. For example, a company might be very likely to succeed if the ensemble says it has a 75% probability of being in the IPO (listed on the stock exchange) or 'acquired by another company' category, while only 25% of its prediction would fall into the failed category.

The researchers trained the models on data sourced from Crunchbase, a crowd-sourced platform containing detailed information on many companies. They married the Crunchbase observations with patent data from the USPTO (United States Patent and Trademark Office). Given the crowd-sourced nature of Crunchbase, it was no surprise to learn that some companies' entries are missing information. This observation

inspired the authors to measure the amount of information missing for each company and use this value as an input to the model. This observation turned out to be one of the most critical features in determining whether a company would be acquired or otherwise fail.

Lead author Greg Ross of Venhound Inc. notes that the ensemble of models, along with novel data features, "generates a level of accuracy, precision and recall that exceeds other similar studies. Investors can use this to quickly evaluate prospects, raise potential red flags and make more informed decisions on the composition of their portfolios."

Provided by KeAi Communications