

## **AAI 695 - Applied Machine Learning**

**Team - Tim Demetriades, Daniel Pelis, Connor DePalma**

### **1) Problem statement - this section defines the machine learning problem.**

Dividend Aristocrats are companies that are in the S&P 500 and have maintained an increasing annual dividend for the past 25 years. These companies are of high importance for the investing strategy called “Dividend snowball”, which relies on the compounding effect caused by reinvesting dividends. The purpose of this “Aristocrat” title is to assure investors this company is a safe investment that will produce reliable dividends for many years. Therefore, we propose a system that will analyze the historical data of current and past Dividend Aristocrats to predict whether a company will lose its title in the coming year.

### **2) Description of data set - this part identifies the source of training data set (provide URL link to the data set if it is from Internet)**

NOBL is an ETF (Exchange Traded Fund) that exclusively invests in S&P 500 Dividend Aristocrats. This means that their holdings are limited to companies that are not only in the S&P 500 but also companies that have been paying out increasing dividends for at least 25 consecutive years. By looking at not only NOBL’s current holdings but also its historical holdings we can get an idea of which companies have been recently added to the list of Dividend Aristocrats along with which companies have been removed. This will allow us to determine if there is a correlation between a Company’s share metrics and its dividend yield and thus in turn its status as a Dividend Aristocrat.  
<https://www.proshares.com/our-etfs/strategic/nobl/>

### **3) Implementation plan - this section briefly describes the tentative plan for implementation, milestones and timeline.**

We will start by creating our data set. We will collect stock information on historical Dividend Aristocrats that have gained, maintained, and lost their status using historical NOBL holdings. Once we have the data we will work through preprocessing our samples. This can include deciding which features we would like our model to pay attention to and attempting to generate new features that could carry significant information. We will then create a model (or collection of models) to train against our collected samples with the goal of predicting whether a company will maintain its Aristocrat status. We will produce useful metrics on our model’s accuracy against test samples as well as areas that could be improved upon in the future.

Our milestones will be: collecting and organizing our dataset, cleaning/preprocessing samples, generating an initial model, creating additional models that could be used to improve results (tentative), and generating model metric after test sample evaluation.

We aim to complete each of the previous milestones in a timely manner. Our current goal is to have a milestone completely every two weeks. This will allow for some adjustments to scheduling if we run into implementation blockers or need more time for a specific milestone.

**4) Team members & task allocation - this section lists names of all team members and defines tasks for each member.**

- Tim
  - Creation of dataset
  - Exploratory data analysis
  - Model research
- Dan
  - Creation of dataset
  - Model optimization / Hyperparameter tuning
  - Model research/implementation
- Connor
  - Creation of dataset
  - Sentiment analysis research on historical news data