AI Applications for Advancing the Study of Behavioral Finance:

A Research Agenda for Predicting IPO Opening Prices using Machine Learning and Unstructured Data

NBER Economics of AI
Big Data and High-Performance Computing for Financial Economics
Cambridge, MA: July 13, 2019

Amanda N. Sweeney

Abstract

The proposed research will explore an Artificial Intelligence (AI) application with potential to benefit the field of Behavioral Finance. This research will investigate how to apply AI tools to help predict the opening prices (driven by supply and demand) of stock after Initial Public Offerings (IPOs). By allowing for the aggregation and analysis of both structured and unstructured data, it may be possible to use AI to eventually develop predictive models that can determine IPO opening price trends. The research will employ a two-phase methodology that culminates with results from a proof-of-concept technical pilot and algorithm.

Scope of Proposed Research

The proposed research will be conducted in two phases. First, in the foundational phase, the challenges to predicting IPO Opening Prices will be analyzed to inform the eventual algorithm / model to be used in the technical pilot. Concurrently, various AI solutions and tools will be researched to confirm their technology readiness levels for such an experimental project. The foundational research of Phase I will enable the planning for a proof-of-concept project during which both structured and unstructured data will be gathered into a data lake, to be used for training an inorganic algorithm that will attempt to predict IPO opening prices in Phase II.

Phase I Research Methodology

The research methodology is as follows:

1. Understand the challenges and gaps in predicting IPO Opening Prices through:
    a. Interviews with PhD candidates and Academics
    b. Journal Review Sample
    c. Historical data collection of IPO offering and opening prices for a sample of no less than 80 companies within the past 10 years

    Outcome: Draft model for functions / variables that affect IPO opening prices

2. Market analysis of AI / inorganic algorithm state of the arts and emerging capabilities of sensors to enable capture and analysis of unstructured data

3. Analysis of research application and development of use cases.

4. Finalize Scope of Research
    a. Refine scope of IPO opening prices research, to be narrowed to specific industry, country, and other control characteristic (to be identified)
    b. Determine data needs and accessibility.
    c. Finalize technology needs for storing data and using unstructured data
        i. Note that current plans include using robotic process automation tools, such as UiPath, with ABBYY for OCR, and other cognitive features to screen scrape and parse social media, news articles, and other sources.

Phase II Research Methodology - Proof of Concept

In phase II, a machine learning / AI proof of concept project will be conducted. To conduct such a project, the author intends to establish partnerships to benefit from additional expertise, capabilities, and data, as described below:

- Expertise: The authors will recruit data science and machine learning PhD candidates to assist with the selection and curation of the data, likely from Carnegie Mellon University. In addition, the said expert will assist with designing the pilot AI solution (algorithms) and iterating to refine the model for the proof of concept project. Further, the author will consult with a privacy expert to be assured that the collection and compilation of the proof of concept data does not violate any known US regulations.
- Technology Capabilities: The author may consider establishing a partnership with a research lab or commercial AI company to leverage data tools, robotic process automation, and cognitive computing platforms. Tools for this research are likely to include:
    - Python and/or R
    - SAS or similar statistical analysis package
    - SQL database
    - UiPath, Blue Prism, or Automation Anywhere for Robotic Process Automation, to be used to collect and vet structured and unstructured data.
    - ABBYY or similar OCR package with cognitive features, to allow for use of unstructured
    - DataRobot, Dynamic AI56, or similar Machine Learning platform to allow for analysis such as sentiment analysis
- Data Partners: The authors will establish partnerships with commercial or academic entities to gain access to statistically significant historical data relevant to the proof of concept project. This data will be anonymized at the counsel of the privacy expert, and supplemented by open source data, such as social media data, to create a limited data lake from which the AI algorithm can draw training data.

Conclusion

The goal of this research is to assess the readiness of AI technology for behavioral finance use cases and establish a framework for using AI in behavioral finance research. As a test case, the research effort will include a foundational phase to refine the use case, which is predicting IPO opening prices. Then, in a second phase, the research will include a technical proof of concept AI project to be conducted with partners in industry and academia. The result of this research will be a repeatable process for refining a behavioral finance research agenda and matching technology to the use case, enabling more experimental research.

About the Author

Amanda Sweeney is the Director of Process Automation at Management Science and Innovation, Inc in Reston, VA. She has more than 13 years experience in technology strategy and implementation. She an expert in the applications of emerging technologies, such as AI, blockchain, and robotics, for various industries including finance, aerospace, and defense. She holds a Bachelor of Science in Economics from

George Mason University in Fairfax, Virginia, USA, as well as a Master of Finance from the Harvard Extension School in Cambridge, Massachusetts, USA. This research agenda represents her interest in applying AI to advance the study of behavioral finance, and a stepping stone to her eventual dissertation.