

# A Natural Language Processing Model for House Price Forecasting

(2018-034)

A machine-learning algorithm that uses text analysis of property description to predict value

## Market Overview

This artificial intelligence (AI) algorithm quantifies the hidden value embedded in property descriptions using natural language processing (NLP) to more accurately forecast property values. The global NLP market is forecasted to be \$16.07 billion in 2021, an increase of \$8.44 billion since 2016. The majority of real estate transaction sites use regression models, which are based on property features such as the number of bedrooms, square footage, construction year, etc. to predict listing prices. Unfortunately, this method excludes one major component of data - the unique qualities of a property that may significantly increase its value. Clemson University researchers have developed a novel tool that accounts for this information, in conjunction with current/past selling prices and the information of comparable properties, to provide economic insights about the impact of real estate uniqueness on sales prices.

## Technical Summary

Real estate agents frequently write property descriptions to convey its unique features and history, reduce market friction, and emphasize any competitive advantages, thereby increasing the likelihood of a sale. This algorithm is able to analyze the text based on mathematical theories and neural network deep learning, an AI model that imitates the human brain, to obtain distributed representations of words, sentences, and paragraphs. This technology preserves the semantic meanings of words within the context of the paragraph and detects more detailed nuances than sentiment analysis methods based on words' positive/negative polarity. Additionally, the algorithm is able to learn the individual agent's writing style and understand abbreviations/typos, as MLS systems impose a 250-words limit on the description length. This technology takes the arrangement order of sentences and paragraphs into consideration, as the relative position of a sentence in a description often implies the importance of the corresponding feature it describes.

### Application

Machine learning, Artificial Intelligence

### Development Stage

Prototype and Animal Studies Complete

### Advantages

- Natural Language Processing (NLP) analyzes hidden value in text descriptions, increasing the property value between 1% and 6%, on average
- Neural network deep learning approach detects detailed nuances, increasing prediction accuracy
- Machine AI is highly detailed, delivering improved market insights to home buyers/sellers, investors, bankers, and policymakers

App Type	Country	Serial No.	Patent No.	CURF Ref. No.	Inventors
NA	NA	NA	NA	2018-034	Dr. Yannan Shen Dr. Yiqiang Han

## About the Inventors



### Dr. Yannan Shen

Assistant Professor of Finance at Clemson University

Dr. Shen is an Assistant Professor of Finance at Clemson University and visiting scholar of the research department at the Federal Reserve Bank of Atlanta. She received her B.S. in Applied Mathematics from the University of California, Berkeley in 2010. She then earned her PhD from Pennsylvania State University in Business Administration in 2016. Since joining Clemson in 2016, her work has been showcased at many finance and economics conferences. Shen is also the single-author winner of the prestigious 2019 American Real Estate Society Manuscript Prize. She is honored to represent Clemson to give research seminars at Cambridge University, the Baruch College, and the Federal Reserve Banks. Due to her efforts, Clemson is the first university that offers machine-learning courses to business undergraduate students. Her current research focuses on financial technology (fintech) and property technology (proptech), in which she applies Machine Learning and Artificial Intelligence to shed light on how technology innovation can help investors and policymakers better monitor financial risks.



### Dr. Yiqiang Han

Research Assistant Professor of Mechanical Engineering at Clemson University

Dr. Yiqiang Han joined the Mechanical Engineering Department at Clemson University as a Research Assistant Professor in 2016. Dr. Han completed his undergraduate studies at the Nanjing University of Aeronautics and Astronautics in 2009. He later finished his M.S. and Ph.D. programs in 2011 and 2015, respectively, at Pennsylvania State University. Before transitioning to Clemson University, Han was a Postdoctoral Research Associate at the Vertical Lift Research Center of Excellence at Penn State. Dr. Han is an active member of AIAA ASE Technical Committee, ASME, ASEE, and VFS. Dr. Han's primary research focuses on distributed optimization, machine learning, and aviation mobility, safety, and navigation. He specializes in data-driven analytical tools to help operate complex cyber-physical systems in a resilient and sustainable manner.

For more  
information on this  
technology contact:

### Andy Bluvas

Technology Commercialization Officer

E: [bluvasa@clemson.edu](mailto:bluvasa@clemson.edu)

P: (864) 656-5157