

FinBERT: Financial Text Mining

Liam Aull and Disha Patel

Table of Contents

01

Introduction

What is FinBERT? Who founded it?

...

02

Background

Key Concepts

...

03

Research Questions

Authors' investigation aim

...

04

Methodology

How the study was conducted?

...

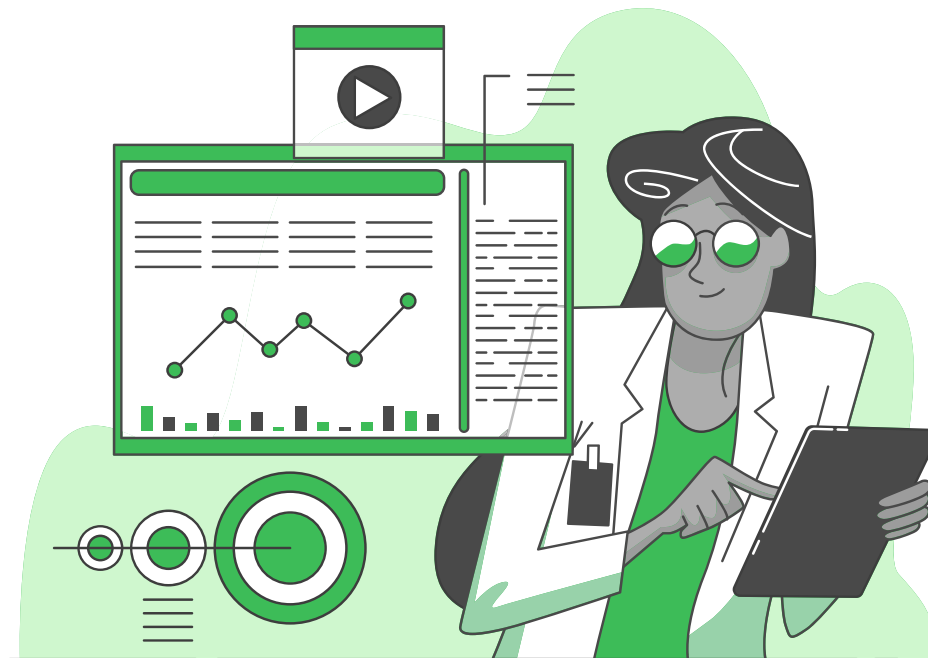


Table of Contents

05

Findings

Main Results

...

06

Commentary

Key Concepts

...



Intro to FinBERT

FinBERT is a pre-trained language model designed specifically for financial text mining.

Importance: It addresses the shortcomings of general models like BERT that struggle with financial terminology

Its founders are: By : Zhuang Liu , Degen Huang, Kaiyu Huang, Zhuang Li and Jun Zhao

The paper was released in 2021




Background: What was FinBERT Before it was FinBERT



Before FinBERT, it was just BERT

BERT is Bidirectional Encoder Representations from Transformers

Prior efforts to improve domain performance (e.x, BioBERT for medical text and more) showed that domain-specific pre-training can outperform generic models.




Building on this Liu et al. propose FinBERT to address the unique linguistic and structural properties of financial texts, such as capitalization patterns, abbreviations, and formal document structures.



Research Questions



The authors aimed to investigate:

1. Can domain-specific pre-training on financial text improve NLP performance over general BERT models?
 2. Do additional self-supervised pre-training tasks (beyond MLM and NSP) enhance model understanding of financial language?
 3. How does FinBERT perform on various financial text-mining applications (e.g., sentiment analysis, question answering, sentence boundary detection)?
- 

Methods and Approach

- FinBERT was created based off of the BERT architecture
 - Trained with 6 self-supervised tasks that better capture financial semantics
 1. Span replace prediction
 2. Capitalization prediction
 3. Token-passage prediction
 4. Sentence deshuffling
 5. Sentence distance prediction
 6. QA relation prediction

Methods and Approach

Tools used :

- BERT's architecture
- TensorFlow
- Horovod
- Pre-trained on combined dataset of general and financial corpora



Findings




Tested on 3 major tasks:

1. Sentence boundary detection
2. Sentiment Analysis
3. Question Answering

FinBERT outperformed BERT in all three tasks for financial texts

Pre-training in financial domain significantly enhances the performance of FinBERT even with smaller training data

Each of the 6 tasks used to train FinBERT contributed positively to the results and performed better than BERT in some testing






Our Thoughts



The paper was boring but the context from which FinBERT is built off of was rather interesting

Not super interested in the financial side of things, but the experiment itself was interesting and its applications seem valuable



Interesting how authors added specifically relevant pre-training tasks to an already existing model (BERT) to serve their purpose



Quiz Time~



1. What is the main reason general BERT performs poorly on financial texts?
 - A. It uses an outdated Transformer model
 - B. It lacks domain-specific vocabulary and semantics
 - C. It can't process numerical data
 - D. It's trained on too much text

2. What does BERT stand for?
 - A. Bidirectional Encoder Representations from Transformers
 - B. Basically, Ernie's Really Trying
 - C. Bidirectional Evaluation of Relative Text
 - D. Building Effective Representations for Transformers



Quiz Time~

3. What makes FinBERT different from the original BERT model?
 - A. It uses extra pre-training tasks designed for financial language.
 - B. It removes all self-supervised training.
 - C. It only trains on movie reviews.
 - D. It replaces words with numbers during training.

4. Which of the following is not one of FinBERT's six pre-training tasks?
 - A. Capitalization Prediction
 - B. Sentence Deshuffling
 - C. Next Sentence Prediction
 - D. Span Replace Prediction

The End



FLIGHTS OF
Imagination

The Muppet Show is a television series that has become a cultural phenomenon. It is a show that has brought joy to millions of children and adults alike. The show is a testament to the power of imagination and the ability of puppets to bring life to the screen. The show is a must-watch for anyone who loves puppetry and television.

Ernie, yellow
Puppet Show, performance from
Bert, green
Puppet Show, performance from

Ernie and Bert, roommates and best buddies on Sesame Street, are reunited in a new, one-act play. Ernie is a playful prankster and Bert is a serious thinker. The two are a great example of the power of friendship and the ability of puppets to bring life to the screen.