Predicting Earnings Calls Surprises Using FinBERT

Sheetal Sudhir Srilatha

Table of Contents

Introduction

I. Introduction

- Project Statement: The goal of this project is to develop a predictive model using FinBERT to identify and forecast earnings call surprises. The project will build on the existing research on sentiment analysis of financial reports using NLP techniques, particularly on the use of FinBERT for analyzing earnings calls.
- Utilized FinBERT (Financial BERT) for my project
  - a pre-trained language model based on BERT (Bidirectional Encoder Representations from Transformers) that has been specifically trained on financial text data
  - trained on a large corpus of financial news articles and financial reports to help it better understand the nuances of financial language

Overview of Earnings Surprises

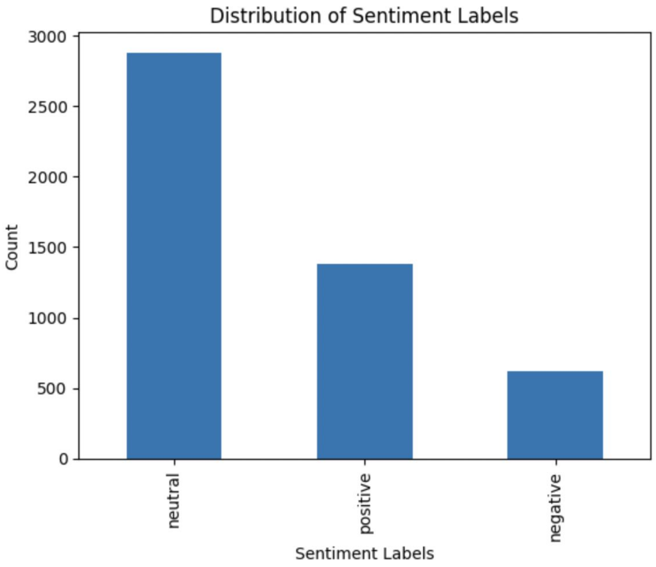
II. Overview of Earnings Surprises

- What are earnings surprises?
  - Earnings call is a conference call held by a public company's executives to discuss their company's financial results for a given reporting period
  - Important metric used in financial analysis
  - Earnings surprise occurs when a company's actual earnings per share (EPS) or revenue exceeds or falls short of the consensus estimate of analysts
- Why is it important to be able predict earnings surprises?
  - Allows investors and traders who want to make informed decisions about buying or selling stocks

Data Collection

III. Data Collection

- Earning call transcripts from 10 different companies were collected, includes:
  - Apple, Advanced Micro Devices, Amazon, ASML, Cisco, Google, Intel, Microsoft, Micron Technology, and Nvidia
- Data source:
  - Dataset: Stock Values and Earnings Call Transcripts: a Sentiment Analysis Dataset — DataverseNL
- Majority class is “neutral” sentiment, used RandomUnderSampler() to undersample the majority class



III. Data Collection – Raw and Processed Data

Raw Data

```
Thomson Reuters StreetEvents Event Transcript
E D I T E D   V E R S I O N

Q4 2015 Intel Corp Earnings Call
JANUARY 14, 2016 / 10:00PM GMT

================================================================================
Corporate Participants
================================================================================

 * Stacy Smith
   Intel Corporation - CFO
 * Mark Henninger
   Intel Corporation - VP of Finance and Director of IR
 * Brian Krzanich
   Intel Corporation - CEO

================================================================================
Conference Call Participiants
================================================================================

 * Vivek Arya
   BofA Merrill Lynch - Analyst
 * Blayne Curtis
   Barclays Capital - Analyst
 * Stacy Rasgon
   Bernstein Research - Analyst
 * Timothy Arcuri
   Cowen and Company - Analyst
 * David Wong
   Wells Fargo Securities, LLC - Analyst
 * Ross Seymore
   Deutsche Bank - Analyst
 * C.J. Muse
   Evercore ISI - Analyst
 * Harlan Sur
   JPMorgan - Analyst
 * Joe Moore
   Morgan Stanley - Analyst
 * John Pitzer
   Credit Suisse - Analyst
 * Chris Danely
   Citigroup - Analyst

================================================================================
Presentation
--------------------------------------------------------------------------------
Operator    [1]
--------------------------------------------------------------------------------

          Good day, ladies and gentlemen, and welcome to the Intel Corporation Q4 2015 earnings conference call.
(Operator Instructions)
As a reminder, this call is being recorded. I would now like to turn the conference over to Mark Henninger. Please go ahead.

--------------------------------------------------------------------------------
Mark Henninger,  Intel Corporation - VP of Finance and Director of IR    [2]
--------------------------------------------------------------------------------

          Thank you, Sabrina, and welcome, everyone, to Intel's fourth-quarter 2015 earnings conference call. By now you should have received a copy of our earnings
release and the CFO commentary that goes along with it. If you've not received both documents they're available on our investor website, INTC.com.
I'm joined today by Brian Krzanich, our CEO, and Stacy Smith, our Chief Financial Officer. In a moment, we'll hear brief remarks from both of them, followed by the
Q&A.
Before we begin, let me remind everyone that today's discussion contains forward-looking statements based on the environment as we currently see it, and as such,
does include risks and uncertainties. Please refer to our press release for more information on the specific risk factors that could cause actual results to differ
materially.
Also, during this call, we'll be using non-GAAP financial measures and references. GAAP financial reconciliations are available in our earnings material, which was
posted on our website, INTC.com, in advance of this call. The forecast that Stacy speaks to today will be on a non-GAAP basis. With that, let me hand it over to
Brian.

--------------------------------------------------------------------------------
Brian Krzanich,  Intel Corporation - CEO    [3]
--------------------------------------------------------------------------------
```

Processed Data

	sentiment	text
0	positive	Revenue for the quarter landed within our guid...
1	positive	Highest quarterly revenue in Co.'s history.\nE...
2	positive	Achieved double-digit growth in:\nGrowth rates...
3	positive	Biggest year ever in most parts of world, with...
4	negative	In 2015, notwithstanding a difficult PC market...

Project Process

IV. Project Process

- Training, Testing Split
  - Used 80-20 split
- Tokenized the text using tokenizer provided by the Hugging Face Transformers library
- Utilized PyTorch's DataLoader to load data in batches during the training and validation phases
- AdamW optimizer is used to optimize the model parameters during training
  - Modified version of the Adam optimizer that uses weight decay to prevent overfitting
  - Learning rate set to 5e-5

IV. Project Process

- FinBERT model is trained for 5 epochs
  - Data is loaded in batches using the data loader
  - Model is trained using the loss calculated from the outputs
  - After training on each batch, the optimizer's gradient is zeroed
- Aforementioned process was completed for both the imbalanced and balanced datasets

Results and Discussions

V. Results and Discussion

- Unbalanced Data:
  - Model achieved a validation accuracy of 0.8567 in the first epoch
  - Reached a peak of 0.9068 in the fifth epoch
  - Training loss decreased consistently over the five epochs, indicates the model was effectively learning from the data.
- Balanced Data
  - Model achieved a validation accuracy of 0.8097 in the first epoch
  - Reached a maximum of 0.8578 in the fifth epoch
  - Training loss was significantly higher in the first epoch and decreased sharply in the subsequent epochs

V. Results and Discussion

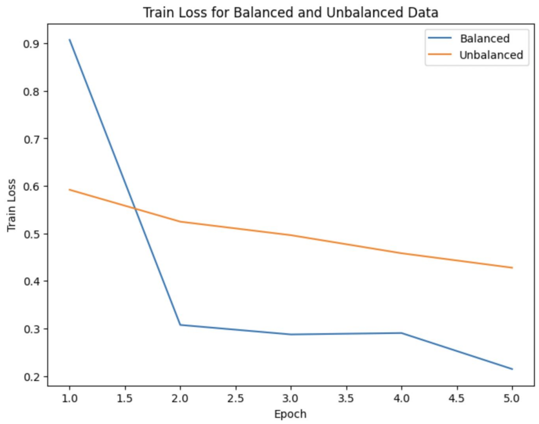


Fig 1. Training loss for balanced vs. unbalanced data

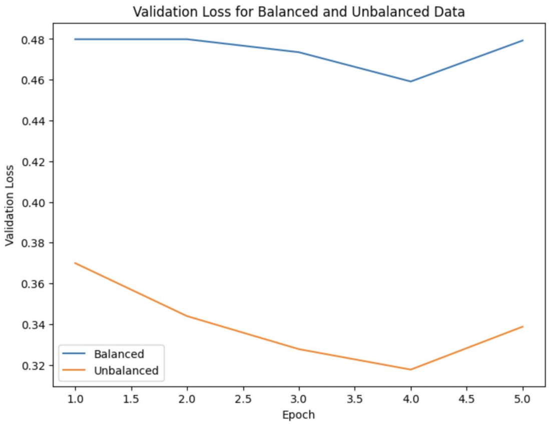


Fig 2. Validation loss for balanced vs. unbalanced data

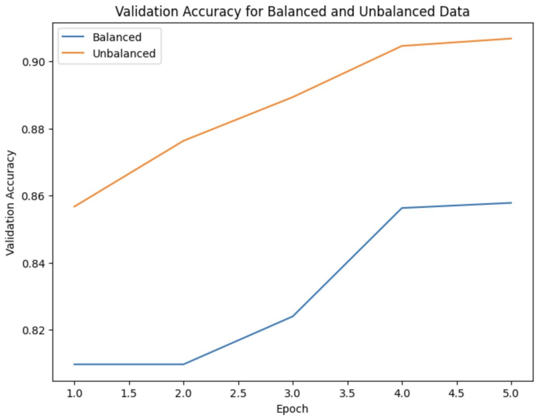


Fig 3. Validation accuracy for balanced vs. unbalanced data

Conclusion and Future Work

VI. Conclusion and Future Work

- FinBERT has some potential in predicting earnings call surprises
- High validation accuracy in both the unbalanced and balanced datasets
  - 90% for the unbalanced data
  - 86% for the balanced data.
- Sharp decrease in training loss in subsequent epochs also suggests that the model was able to learn from the data and improve its performance over time

VI. Conclusion and Future Work

- Limitations
  - Sensitivity to imbalanced data:
    - In the unbalanced dataset, model had a tendency to predict neutral even when the actual label was positive or negative.
  - FinBERT is limited to the English language
    - may not perform as well on non-English earnings calls
  - Small number of companies representing this dataset
    - limited generalizability of the model to other companies
    - include data from more companies in the database

References

- Li, W., & Ye, X. (2021). Predicting the market response to corporate earnings calls using FinBERT. International Journal of Information Management, 56, 102222.
- Hamid, F., Masud, M. M., & Al Hasan, M. (2021). FinBERT-Cap: Capitalizing financial domain knowledge for earnings surprise prediction. Expert Systems with Applications, 175, 114819.
- Shvets, M., & Kotlyarov, I. (2021). Predicting Earnings Surprises: A Machine Learning Approach. arXiv preprint arXiv:2103.13189.