A transformer-based model for default prediction in mid-cap corporate markets

KAMESH KORANGI, CHRISTOPHE MUES, CRISTIÁN BRAVO

CBRAVORO@UWO.CA

@CBRAVOR

The Banking Analytics Lab

One of the largest groups in the world researching analytics in banking, with a focus on credit risk.

- 3 postdoctoral researchers.
- 12 PhD students across three continents.
- A large network of collaborators all around the world.

Recognized worldwide!

- Markov et al. (2022):
  - Top 2 researcher in the area of credit scoring. Only in two lists (also in Louzada et al., 2017).
  - Two works (Bravo et al., 2013 and Verbraken et al., 2014) are now industry standards.

Visit our website!

https://thebal.ai

SCAN ME

Agenda

- Mid-caps
- Problem
- Relevant Literature
- Contributions
- Data Description
- Results
- Conclusion

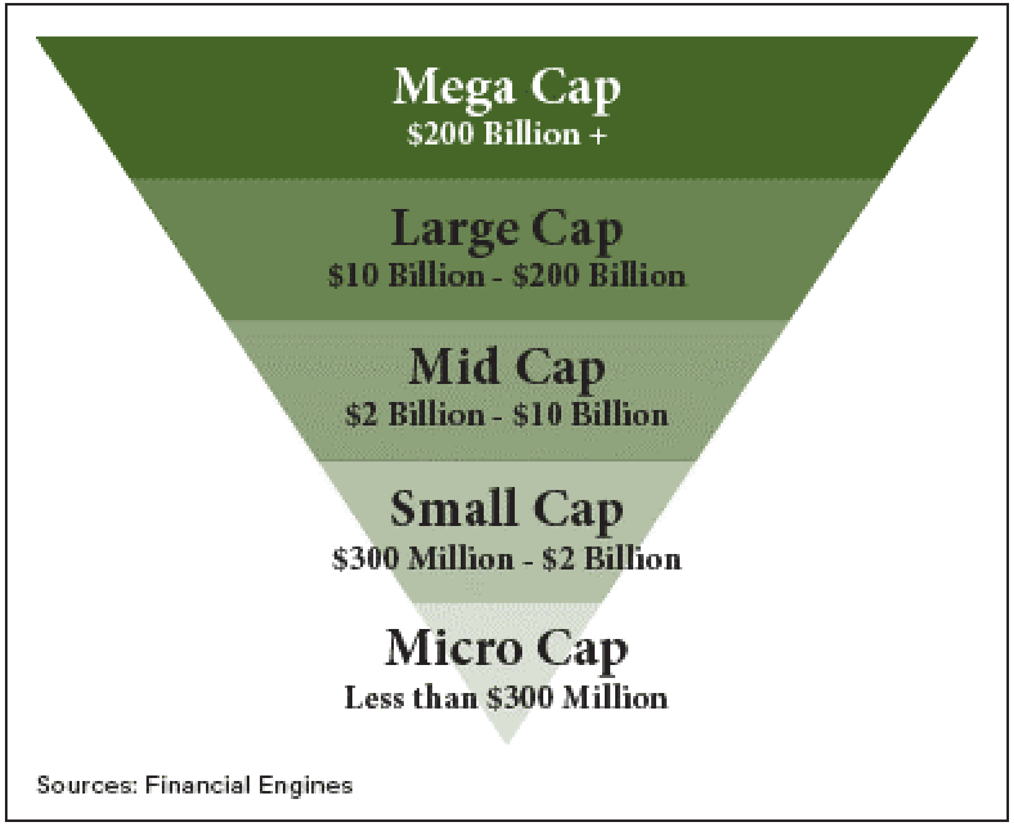
Mid-caps

Mid-cap companies are publicly traded companies with less than US$10 billion market capitalisation

Their debt has a shorter maturity period than large-caps

Typically hold a non-investment grade credit rating, implying higher credit risk

Lots of disaggregated information, more dependant on general economic conditions.



Problem

Mid-cap challenges

- Credit spreads or prices implied by the models often underestimate from what is empirically observed[1], and is more pronounced for mid-caps[2]
- This mispricing of risk lead to unexpected losses for lenders
- Credit risk is not easily separable from market risk, especially for mid-cap companies[3]

Modelling challenges

- Ability to use different kinds of data and extract relationships
- Time horizon of default prediction models[4]

In the Literature…

Three types of credit risk models in this space – statistical models[5], structural models[6] and reduced form models[7].

Machine learning models shown to improve performance, particularly ensemble of models in default prediction[8,9]

Deep learning models have produced state-of-the-art results in other domains and been applied with textual data for corporates[10] and for SMEs[11] with promising results

Our proposal

We construct an **ad-hoc transformer model**[12] -- state-of-the art in natural language processing - for **time series panel data with correlated outputs**

General Aim and Paper Contributions I

We develop a novel approach to credit risk modelling using deep learning in midcaps

- first to propose a **transformer-encoder model for corporate default risk** modelling
- a **framework for multi-modal learning** that can combine the different data sources and allows for a differential training approach

Single combined model for term structure of default probabilities

- Probabilities are obtained at 3m, 6m, 1y, 2y and 3y **simultaneously**.

General Aim and Paper Contributions II

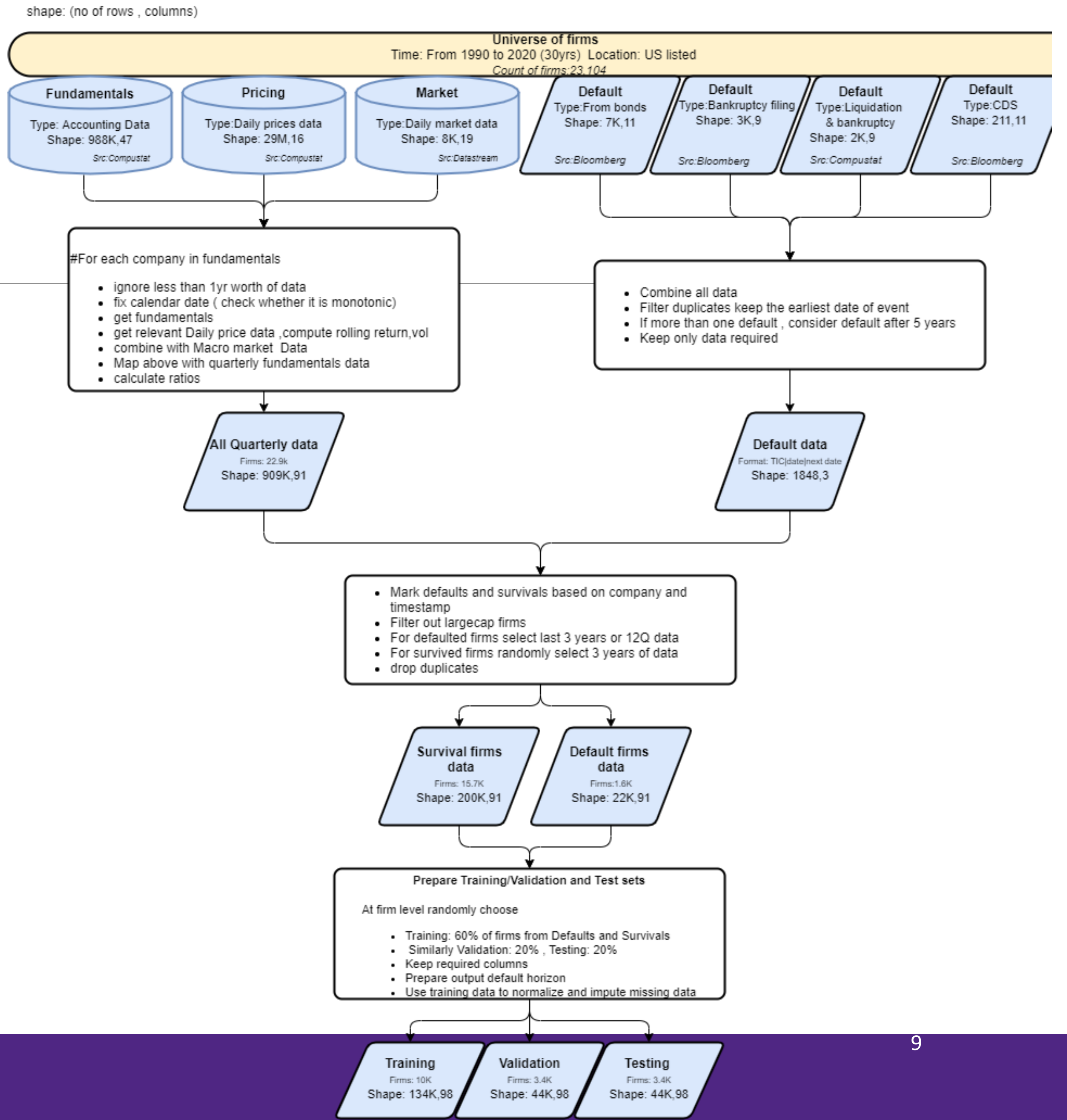
We also make the model interpretable

- Utilize **attention heatmap visualizations** to show the model learning between defaults and survivals

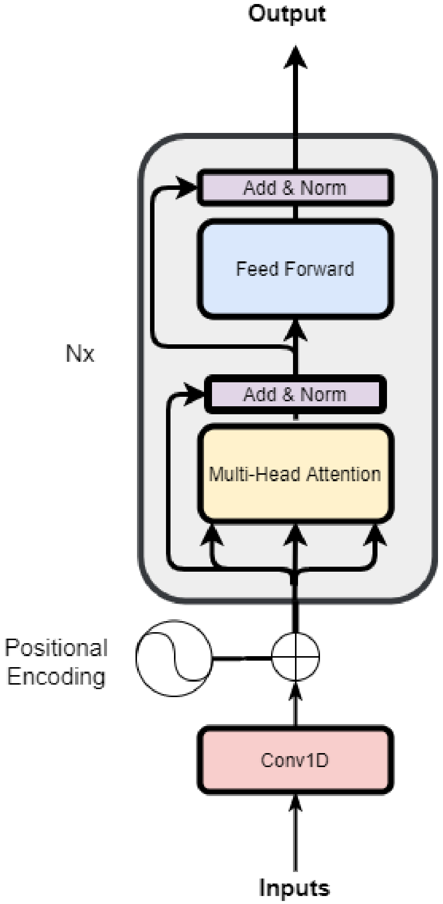
- Shapley values to quantify the relative importance of **groups of variables** to answer
  - Which data sources are important
  - Which time period is important

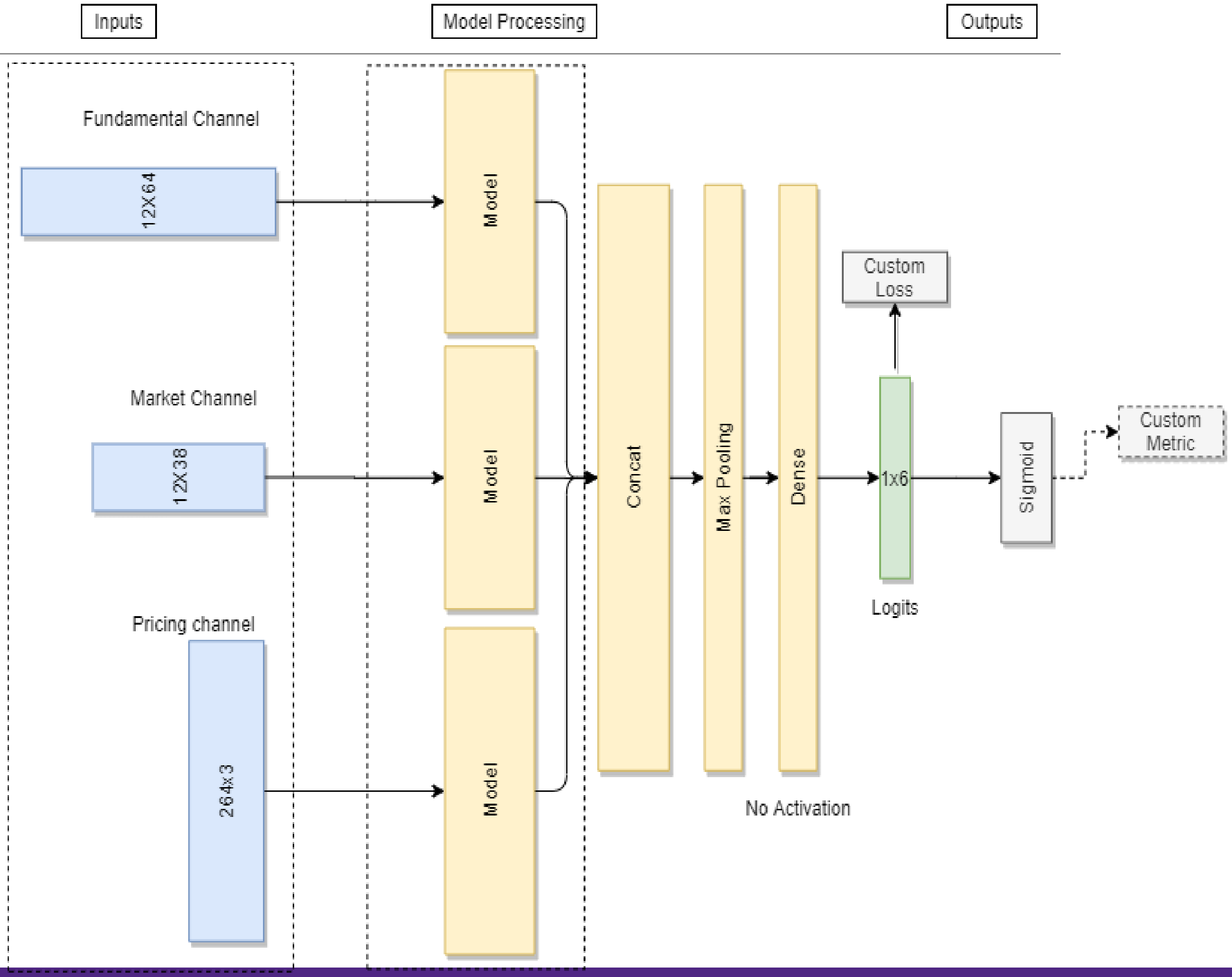
Data description

- 30 years of data related to mid-cap companies listed in the US from 1990 to 2020
- Three different data sources
- Wider definition of default



Model architecture





Transformer Encoder model

Multi-modal architecture

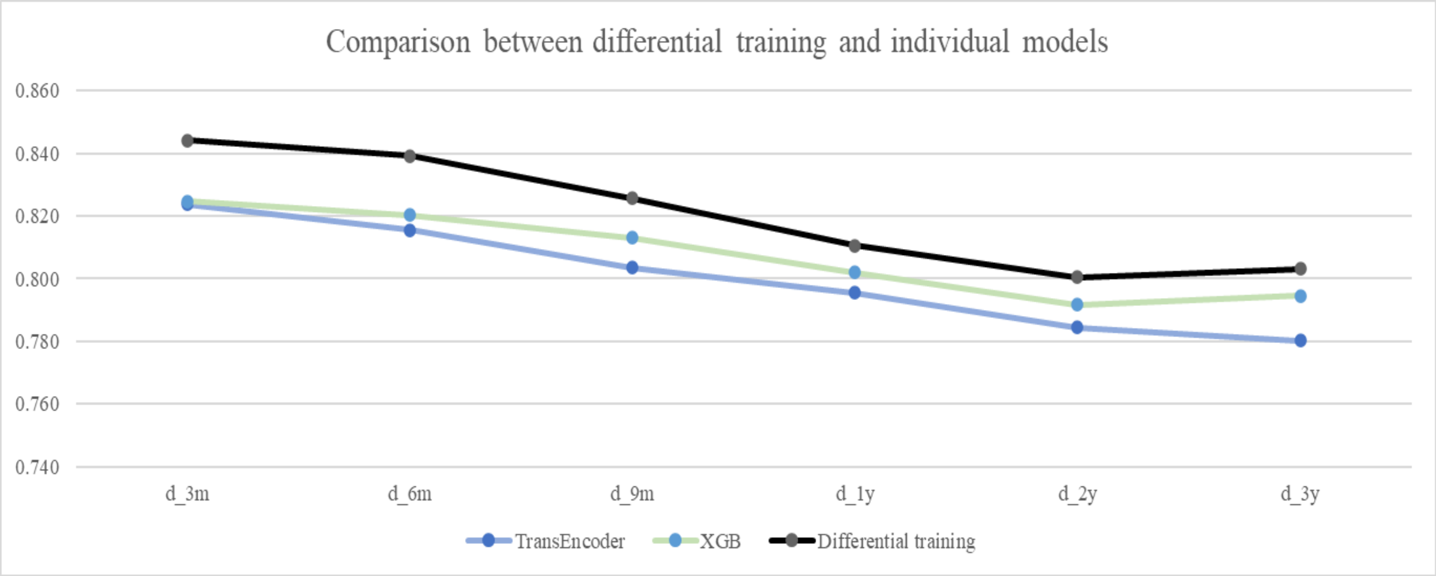
Results

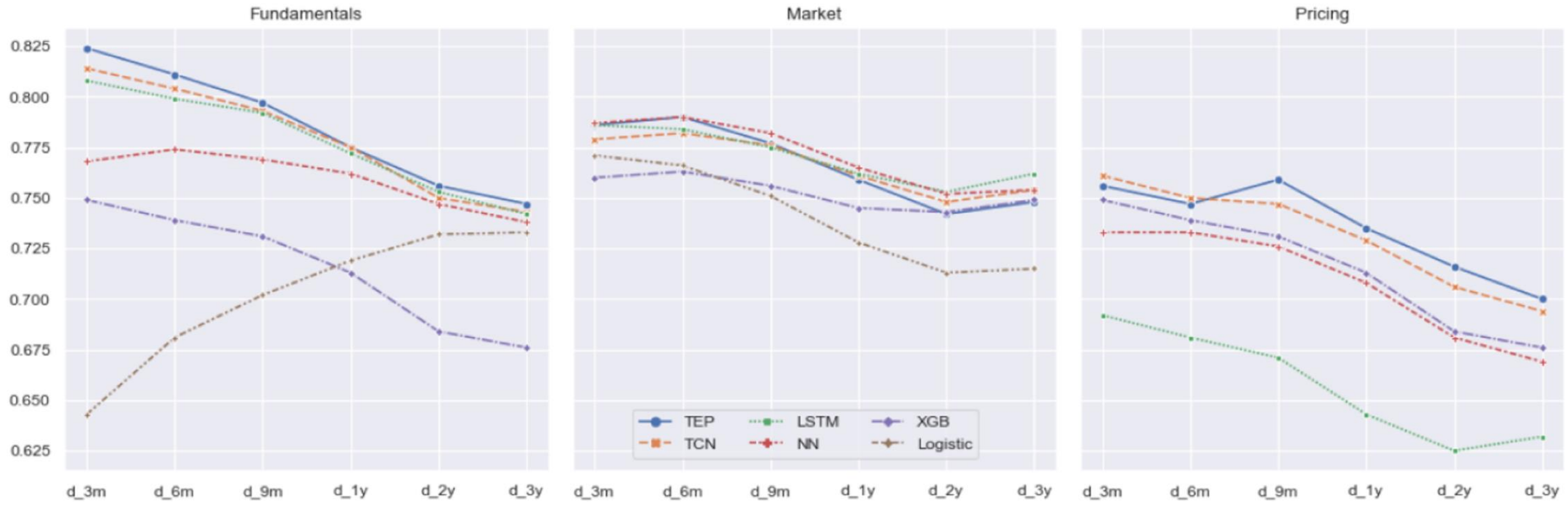
Deep learning models perform better with individual data

Long-range dependencies could be found with TransEncoder and TCN as can be seen from daily pricing data channel

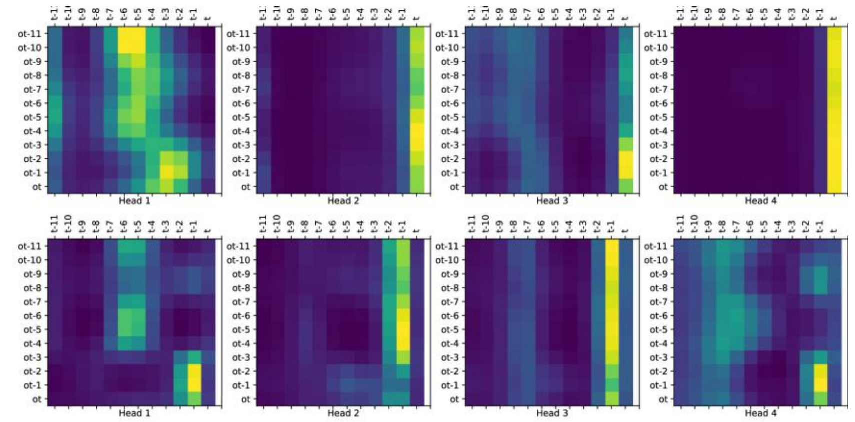
With all data, **multi-modal architecture allows deep learning models to outperform XGB**

AUC@ROC Different data sources	Fundamentals only	Market data	Pricing data
TransEncoder	**0.785**	0.767	**0.736**
TCN	**0.780**	0.767	**0.731**
LSTM	0.777	**0.770**	0.657
NN	0.756	**0.772**	0.708
XGB	0.715	0.752	0.715
Logistic	0.702	0.741	0.535

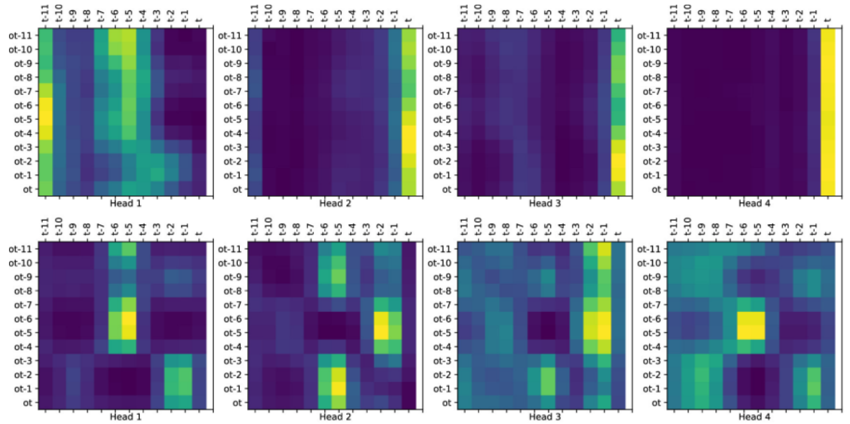



Fundamentals
Market
Pricing
0.825
0.800
0.775
0.750
0.725
0.700
0.675
0.650
0.625
d_3m
d_6m
d_9m
d_1y
d_2y
d_3y
TEP
TCN
LSTM
NN
XGB
Logistic

Input: Quarterly and daily data channnels		**AUC**						
Method	**Regime**	**Average**	**d_3m**	**d_6m**	**d_9m**	**d_1y**	**d_2y**	**d_3y**
TCN								
Training together	2	0.812	0.829	0.821	0.813	0.808	0.802	0.799
Pricing channel freeze	3	0.817	0.828	0.833	0.822	0.812	0.805	0.802
Market and pricing channel freeze	3	0.821	0.839	0.814	0.820	0.814	0.826	0.812
TEP								
Training together	2	0.835	0.858	0.848	0.843	0.832	0.820	0.812
Pricing channel freeze	3	0.841	0.860	0.852	0.846	0.841	0.824	0.822
Market and pricing channel freeze	3	**0.847**	0.867	0.860	0.850	0.847	0.833	0.824



(b) Average weights for firms with no observed default



(a) Average weights for firms that default

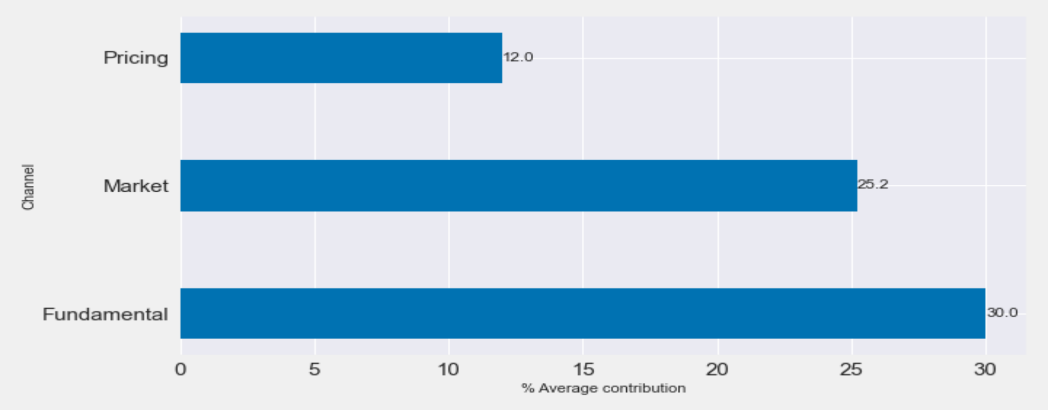
Interpreting the results

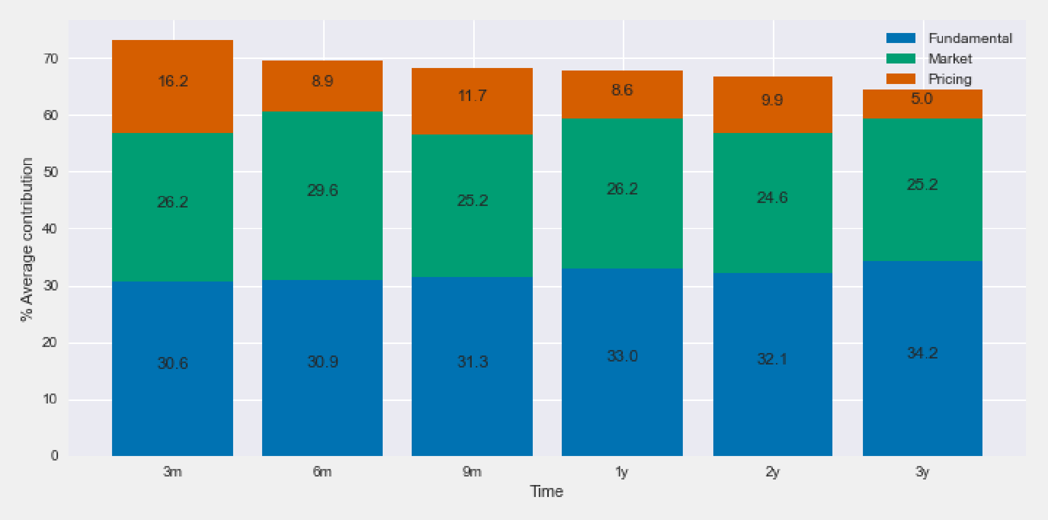
- Difference in learning between the default firms and non-default firms
- Each head focuses on different aspect of the input
- Input last 12 quarters of data mapped to output representation
- Higher weights colored yellow

Interpretation of the Results - Shapley

- **Accounting information is most important data source for default prediction** in mid-caps accounting for 30% importance
- **Macro economic environment more important over medium term than equity performance of the company**
- **Pricing channel provides signalling in the short term**
- Temporally , present data is more important for prediction over 52% of performance from present accounting information compared to 12.4% from past two years.

	Shapley values	
Channel	Present 1 year	Past 2 years
Fundamental	52.3	12.4
Market	35.1	20.0
Pricing	38.4	9.3





Conclusion

Deep learning models improve predictive power especially with complex models and multimodal architectures

Term structure of probabilities produced within single model by manipulating the objective function

Provide interpretability to the results by using heatmaps and custom methodology for groups of variables based on Shapley values

Could extend models with unstructured data like text and audio data

Multi modal architecture could be extended to produce new kind of scorecard models for credit risk

Read the paper! Korangi et al. (2023) @ EJOR

SCAN ME

Banking Analytics Lab

Western
UNIVERSITY · CANADA

UNIVERSITY OF Southampton

Thank you!

Q&A

References

1. Amato, J.D. and Remolona, E.M., 2003. The credit spread puzzle. *BIS Quarterly Review, December*.
2. Feldhütter, P. and Schaefer, S.M., 2018. The myth of the credit spread puzzle. *The Review of Financial Studies*
3. Jarrow, R.A. and Turnbull, S.M., 2000. The intersection of market and credit risk. Journal of Banking & Finance
4. du Jardin, P. and Séverin, E., 2012. Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time. *European Journal of Operational Research*
5. Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, *23*(4), pp.589-609.
6. Merton, R.C., 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, *29*(2), pp.449-470.
7. Jarrow, R.A. and Turnbull, S.M., 1995. Pricing derivatives on financial securities subject to credit risk. *The journal of finance*, *50*(1), pp.53-85.
8. Alaka, H.A., Oyedele, L.O., Owolabi, H.A., Kumar, V., Ajayi, S.O., Akinade, O.O. and Bilal, M., 2018. Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*
9. Barboza, F., Kimura, H. and Altman, E., 2017. Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, pp.405-417.
10. Mai, F., Tian, S., Lee, C. and Ma, L., 2019. Deep learning models for bankruptcy prediction using textual disclosures. *European journal of operational research*, *274*(2), pp.743-758.
11. Stevenson, M., Mues, C. and Bravo, C., 2021. The value of text for small business default prediction: A deep learning approach. *European Journal of Operational Research*.
12. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I., 2017. Attention is all you need. *arXiv preprint arXiv:1706.03762*.