

# How are attention and sentiment scores derived?

February 12, 2025

## 1 General purpose

Starting from the big picture:

- 10-K and 10-Q filings contain many paragraphs;
- I identify the general accounting theme of each paragraph in a 10-K/Q filing;
- The number of paragraphs with topic  $T_j$  as the main theme, normalized by the total number of paragraphs of the corresponding filing, is the attention score of  $T_j$  of the filing;
- How to identify the theme of a paragraph?

## 2 Technical details

- I follow the approach of [Burke et al. \(2023\)](#). In particular, a language model is trained in a supervised learning manner to classify paragraphs in 10-K and 10-Q filings into 31 accounting topics;
- Due to the supervised learning nature of the task, a dataset with accounting topics as the labels is required. Accordingly, I rely on the Financial Statement and Notes (FSN) dataset published by the US Securities and Exchange Commission (SEC).<sup>1</sup>

---

<sup>1</sup>Data available at: <https://www.sec.gov/data-research/financial-statement-notes-data-sets>

This data set consists of enormous number of notes extracted from various firm disclosures (e.g., 10-K, 10-Q, 10-K/Q, 6-K, 8-K, 20-F, etc.). These notes are associated with eXtensible Business Reporting Language (XBRL) tags. There are 84,293 XBRL tags in my training dataset;

- [Burke et al. \(2023\)](#) select 85 XBRL tags covering 92.5% of all taxonomy tags in 10-K filings. Subsequently, they assign these 85 XBRL tags to 31 accounting topics from Financial Accounting Standards Board (FASB) Codification.<sup>2</sup> These accounting topics include: *Asset Retirement; Business Conditions; Cash; Collaborative Arrangements; Commitments and Contingencies; Compensation; Consolidation; Debt; Deferred Revenue; Derivatives and Hedging; Equity; Fair Value; Financial Services; Goodwill and Intangibles; Insurance; Inventory; Investment; Leases; Liabilities; Oil and Gas; Pension and Post-employment; Property, Plant, and Equipment (PPE); Real Estate; Receivables; Related Party; Research and Development (R&D); Restructuring; Revenue; Tax; Transfers and Servicing; and Warranties and Guarantees*. The detailed list of the accounting topics and the corresponding XBRL tags are given in Appendix A;
- I use the FSN dataset and exclude all notes from 10-K/Q filings from the training dataset. This helps us to avoid look-ahead bias as we want to derive the attention and sentiment of the accounting topics in those filings;
- I map [Burke et al. \(2023\)](#)’s 85 XBRL tags into the 31 accounting topics. The other tags are labeled as ‘Others’;
- Due to the large number of the ‘Others’ label compared to the others in the FSN dataset, I sample 1% of notes associated with this label. In the end, the final data set includes 350,258 notes associated with the 31 accounting labels and one ‘Others’ label;
- I randomly split this data into three parts: training, validating, and testing parts with the 50:25:25 ratio, respectively;
- I train FinBERT ([Huang et al., 2023](#)) using the constructed dataset to classify notes into the 32 accounting topics. The validation F1 score is 96.34%;
- I then use this customized FinBERT to classify all paragraphs in all 10-K/Q filings into 31 accounting topics;

---

<sup>2</sup>I exclude the topic ‘Other Assets’ of [Burke et al. \(2023\)](#) due to the opaque contents it covers.

- I use the original FinBERT to score the sentiment class for each paragraph those are assigned to one of the 31 accounting topics. I neglect those which are assigned to ‘Others’.

## References

Burke, J., Hoitash, R., Hoitash, U., & Xiao, S. X. (2023). Using a large language model for accounting topic classification. *Available at SSRN 4484489*.

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 40(2), 806–841.

## A Accounting topics from Burke et al. (2023)

Accounting Topics	XBRL TextBlock Taxonomy Tags
<i>Asset Retirement</i>	AssetRetirementObligationDisclosureTextBlock
<i>Business Combinations</i>	BusinessCombinationDisclosureTextBlock MergersAcquisitionsAndDispositionsDisclosuresTextBlock BusinessAcquisitionIntegrationRestructuringAndOtherRelatedCostsTextBlock
<i>Cash</i>	CashAndCashEquivalentsDisclosureTextBlock CashCashEquivalentsAndMarketableSecuritiesTextBlock CashCashEquivalentsAndShortTermInvestmentsTextBlock
<i>Collaborative Arrangements</i>	CollaborativeArrangementDisclosureTextBlock
<i>Commitments and Contingencies</i>	CommitmentsDisclosureTextBlock CommitmentsAndContingenciesDisclosureTextBlock CommitmentsContingenciesAndGuaranteesTextBlock LegalMattersAndContingenciesTextBlock LossContingencyDisclosures
<i>Compensation</i>	CompensationAndEmployeeBenefitPlansTextBlock CompensationRelatedCostsGeneralTextBlock CompensationAndEmployeeBenefitPlansOtherThanShareBasedCompensationTextBlock DisclosureOfCompensationRelatedCostsShareBasedPaymentsTextBlock ShareholdersEquityAndShareBasedPaymentsTextBlock DisclosureOfShareBasedCompensationArrangementsByShareBasedPaymentAwardTextBlock
<i>Consolidation</i>	MinorityInterestDisclosureTextBlock VariableInterestEntityDisclosureTextBlock
<i>Debt</i>	DebtDisclosureTextBlock LongTermDebtTextBlock ShortTermDebtTextBlock FederalHomeLoanBankAdvancesDisclosureTextBlock SubordinatedBorrowingsDisclosureTextBlock
<i>Deferred Revenue</i>	DeferredRevenueDisclosureTextBlock

<i>Derivatives and Hedging</i>	DerivativeInstrumentsAndHedgingActivitiesDisclosureTextBlock FinancialInstrumentsDisclosureTextBlock DerivativesAndFairValueTextBlock
<i>Equity</i>	StockholdersEquityNoteDisclosureTextBlock PreferredStockTextBlock TreasuryStockTextBlock PartnersCapitalNotesDisclosureTextBlock
<i>Fair Value</i>	FairValueDisclosuresTextBlock FairValueMeasurementInputsDisclosureTextBlock
<i>Financial Services</i>	RegulatoryCapitalRequirementsUnderBankingRegulationsTextBlock
<i>Goodwill and Intangibles</i>	GoodwillAndIntangibleAssetsDisclosureTextBlock GoodwillDisclosureTextBlock IntangibleAssetsDisclosureTextBlock
<i>Insurance</i>	ReinsuranceTextBlock
<i>Inventory</i>	InventoryDisclosureTextBlock
<i>Investments</i>	InvestmentsInDebtAndMarketableEquitySecuritiesAndCertainTradingAssetsDisclosureTextBlock InvestmentTextBlock InvestmentHoldingsTextBlock InvestmentsAndOtherNoncurrentAssetsTextBlock MarketableSecuritiesTextBlock EquityMethodInvestmentsDisclosureTextBlock
<i>Leases</i>	LeasesOfLesseeDisclosureTextBlock DebtAndCapitalLeasesDisclosuresTextBlock LeasesOfLessorDisclosureTextBlock OperatingLeasesOfLesseeDisclosureTextBlock CapitalLeasesInFinancialStatementsOfLesseeDisclosureTextBlock LesseeOperatingLeasesTextBlock OperatingLeasesOfLessorDisclosureTextBlock LesseeFinanceLeasesTextBlock
<i>Liabilities</i>	AccountsPayableAndAccruedLiabilitiesDisclosureTextBlock DepositLiabilitiesDisclosuresTextBlock OtherLiabilitiesDisclosureTextBlock AccountsPayableAccruedLiabilitiesAndOtherLiabilitiesDisclosureCurrentTextBlock
<i>Oil and Gas</i>	OilAndGasExplorationAndProductionIndustriesDisclosuresTextBlock
<i>Pension and Post-employment</i>	PensionAndOtherPostretirementBenefitsDisclosureTextBlock PostemploymentBenefitsDisclosureTextBlock
<i>Property, Plant, and Equipment</i>	PropertyPlantAndEquipmentDisclosureTextBlock DisposalGroupsIncludingDiscontinuedOperationsDisclosureTextBlock
<i>Real Estate</i>	RealEstateDisclosureTextBlock RealEstateOwnedTextBlock RealEstateAndAccumulatedDepreciationDisclosureTextBlock
<i>Receivables</i>	LoansNotesTradeAndOtherReceivablesDisclosureTextBlock FinancingReceivablesTextBlock AllowanceForCreditLossesTextBlock
<i>Related Party</i>	RelatedPartyTransactionsDisclosureTextBlock
<i>Research and Development</i>	ResearchDevelopmentAndComputerSoftwareDisclosureTextBlock
<i>Restructuring</i>	AssetImpairmentChargesTextBlock RestructuringImpairmentAndOtherActivitiesDisclosureTextBlock RestructuringAndRelatedActivitiesDisclosureTextBlock
<i>Revenue</i>	LongTermContractsOrProgramsDisclosureTextBlock

	RevenueFromContractWithCustomerTextBlock
<i>Tax</i>	IncomeTaxDisclosureTextBlock
<i>Transfers and Servicing</i>	TransfersAndServicingOfFinancialAssetsTextBlock RepurchaseAgreementsResaleAgreementsSecuritiesBorrowedAndSecuritiesLoanedDisclosureTextBlock
<i>Warranties and Guarantees</i>	ProductWarrantyDisclosureTextBlock GuaranteesTextBlock

**Table 1:** Accounting topics