POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Uncovering latent themes and emotional dynamics in youth precariousness

Tesi di Laurea Magistrale in
Mathematical Engineering - Statistical Learning

Francesco Tomasi, 252533

Abstract: This thesis investigates the subjective experience of precariousness among Spanish youth by applying advanced Natural Language Processing techniques to open-ended survey responses from the *VulnYouth* project. While traditional research often defines precariousness through objective labour market indicators, this study aims to uncover the latent themes and underlying emotions that young people use to conceptualize their own condition.
A primary challenge addressed by this work is the detection of fine-grained differences between generally negative emotions, overcoming the limitations of traditional literature which largely focuses on binary positive-versus-negative emotion detection.
Methodologically, the research integrates Structural Topic Modeling to map the thematic structure of the discourse with Transformer-based Large Language Models—specifically RoBERTa fine-tuned on GoEmotions for categorical classification and BERT trained on EmoBank for dimensional analysis (Valence-Arousal-Dominance).
The results identify 20 latent topics that portray precariousness, and the emotional analysis reveals a landscape dominated by disappointment and sadness with a notable absence of high-arousal emotions like anger. Another sociological finding is the neutrality paradox: respondents with the highest levels of objective material deprivation are significantly more likely to use neutral, detached language, suggesting a mechanism of resignation. This thesis concludes that youth precariousness in Spain is characterized by a profound sense of disappointed resignation and a perceived loss of agency.

Advisor:
Alessandra Guglielmi

Co-advisors:
Lara Maestripieri

Academic year:
2024-2025

Key-words: youth precariousness; emotion detection, Structural Topic Modeling; BERT; RoBERTa

Contents

1. Introduction

Precariousness has emerged in recent decades as a structural and significant element of contemporary society, attracting increasing attention from social scientists, policymakers, and public institutions. In the past, precariousness has been analyzed primarily in relation to the labour market, while today it is increasingly recognized as a multidimensional phenomenon, with implications that go beyond the economic sphere and more broadly concern living conditions and individual well-being (Vosko et al., 2009).

In this context, a growing number of studies have highlighted how occupational instability can have significant consequences on mental health, particularly among younger generations, who are more exposed to temporary contracts and job uncertainty. Vikram Patel et al. (2007) identified the decline in mental health among young people as a global public health challenge, while Joan Benach et al. (2023) emphasize the need to investigate the structural and social factors that contribute to the deterioration of psychological well-being in this population group.

From this perspective, understanding the relationship between job insecurity and mental health becomes crucial. As highlighted by Irvine and Rose (2022), for younger people job insecurity represents not only an economic condition, but also a subjective experience characterized by uncertainty, vulnerability and loss of control, which can profoundly influence psychological well-being. Analyzing these dynamics is therefore essential to develop targeted policies and interventions capable of mitigating the negative effects of precariousness and promoting more stable and sustainable living conditions for young people.

These dynamics are particularly visible in Spain. As noted by Bolíbar et al. (2019), Spain is characterized by a dual labour market with an exceptionally high share of temporary employment in Europe, especially among young workers. This institutional configuration produces cumulative disadvantage, limiting the accumulation of economic and social capital necessary for a stable adult life. Understanding how precariousness is subjectively defined and emotionally experienced by young people in this context is therefore essential for designing labour market policies, welfare measures, and mental health interventions.

This thesis contributes to this debate by analyzing open-ended responses from the *VulnYouth* project (IGOP, Institut de Govern i Polítiques Públiques, 2023), a study funded by the La Caixa Foundation that examines intersectional vulnerabilities and the link between mental health and job insecurity among young people in Spain. Specifically, we focus on the textual answers to the question: *"What does precariousness mean to you?"* collected from Spaniards aged 20 to 34. Unlike predefined survey choices, these open-ended responses allow for a nuanced understanding of the phenomenon, capturing the emotional and cognitive burden of instability that standard metrics often miss.

To analyze this unstructured text, this thesis adopts a dual computational strategy. First, *Topic Modeling* is used to uncover latent thematic structures within the responses. Through Structural Topic Modeling (STM), we identify how precariousness is cognitively framed—whether primarily as material deprivation (e.g., wages, housing), institutional exclusion (e.g., lack of rights), or future uncertainty—and how these patterns vary across socio-demographic covariates, such as gender or migratory background.

Second, *Emotion Detection* is employed to capture the affective dimension of these narratives. While topic modeling identifies *what* respondents talk about, emotion analysis reveals *how* they feel about it. Using transformer-based architectures derived from Google's BERT (Devlin et al., 2018), each response is analyzed through two complementary approaches: a fine-grained categorical classification to detect discrete emotions (e.g., anger, sadness) and a dimensional assessment to quantify Valence, Arousal, and Dominance scores. By focusing on the narratives and emotional experiences of young individuals, this thesis contributes to sociological debates on youth transitions and the impact of precariousness on mental health, moving beyond purely objective indicators of employment status.

The work is structured as follows. Section 2 discusses the State of the Art regarding the application of text analysis in the social sciences and the specific challenges of modeling short, non-English texts. Section 3 describes the VulnYouth dataset and the pre-processing pipelines. Section 4 details the methods, providing the technical description of the STM and BERT-based classifiers. Section 5 details the model configurations, and Section 6 presents the results. Finally, Section 7 provides the conclusions.

2. State of the art

2.1. Text analysis for the social sciences

In recent years, computational text analysis has become an increasingly important tool for investigating subjective experiences and collective representations. Traditionally, analyzing open-ended survey responses in sociology has relied on *manual annotation*. Human coders interpret text to assign themes or sentiment labels. While this approach captures linguistic nuances like irony or sarcasm, it is costly, time-consuming, and hard to scale (Mohammad, 2016). Furthermore, human coding is subjective; even trained annotators often struggle with consistent labeling when faced with complex, mixed emotions. Consequently, social science research has increasingly turned to computational methods—specifically Natural Language Processing (NLP)—to uncover latent themes and affective states at scale.

Applying these computational methods to the VulnYouth dataset presents two specific challenges that define the current state of the art in this domain:

- Most state-of-the-art NLP models (such as standard BERT and RoBERTa) are pre-trained primarily on English corpora. While multilingual models exist, they often lag behind their English counterparts in capturing subtle emotional nuances. To address this, this study opts to translate the original Spanish responses into English. This approach allows us to leverage the full capability of English-native models without the performance degradation associated with zero-shot cross-lingual transfer.
- The short-text problem: The data consists of open-ended survey responses, which are typically very brief. This is a known difficulty in NLP, as traditional "bag-of-words" models often fail to capture context in short segments. This necessitates the use of Transformer-based models, which rely on "attention mechanisms" to understand the contextual relationship between words even in limited sentences (Jurafsky et al., 2024).

2.2. Topic modeling

To analyze the *content* of these translated responses, we employ *Topic Modeling*. While emotion detection captures *how* respondents feel about precariousness, topic modeling identifies *what* they refer to when expressing those feelings—such as housing conditions, low wages, job instability, or limited future prospects.

Topic modeling comprises a class of unsupervised statistical techniques designed to uncover latent thematic structures within a corpus of documents. The standard approach, *Latent Dirichlet Allocation (LDA)*, models documents as mixtures of topics but treats them as independent observations, thereby overlooking the social context in which texts are produced.

To overcome this limitation, we employ the *Structural Topic Model (STM)*, which extends LDA by incorporating document-level metadata directly into the model. By including covariates such as employment vulnerability, job insecurity, economic strain, and economic independence (from the VulnYouth dataset), STM allows us to estimate not only the thematic structure of the responses but also how topic prevalence varies across different social positions.

This approach enables us to link distinct emotional profiles (e.g., sadness versus disappointment) to specific dimensions of precariousness, connecting affective expression to underlying socio-economic conditions.

2.3. Emotion detection

While topic analysis successfully maps the thematic landscape of the responses, capturing the full subjective experience requires an additional layer of analysis. To bridge the gap between the identified themes and the respondents' inner state, we integrate *Emotion Detection*. This study adopts a dual approach to emotion analysis, aligning with the primary emotion modeling techniques identified in the literature (see Al Maruf et al, 2024). We list two such techniques:

- Categorical emotion model: This approach relies on the identification of distinct, separate emotional states. We operationalize this using the *GoEmotions* dataset, which categorizes text into a fine-grained taxonomy of 28 distinct emotional labels.
- Dimensional emotion model: This approach characterizes emotions as continuous coordinates within a vector space. Specifically, we utilize the VAD framework, which defines emotional states through *Valence* (the degree of pleasantness or unpleasantness), *Arousal* (the intensity of the emotion, ranging from high activation like anger to low activation like calm or resignation), and *Dominance* (the degree of control or agency perceived in the situation).

In the next two subsections, we will go deeper into the two methods.

2.3.1 Categorical emotion model

To operationalize the categorical approach, this research utilizes a method based on RoBERTa (A Robustly Optimized BERT Pretraining Approach).

This choice is motivated by the model's underlying *Transformer* architecture, which utilizes a *Self-Attention* mechanism. Unlike traditional models that process text sequentially, Transformers analyze the entire sequence simultaneously, calculating the relationship of every word to every other word. This capability is particularly crucial for the VulnYouth dataset: it allows the model to resolve ambiguities in short, open-ended responses by leveraging the specific context of each word.

The primary objective of this method is to perform multi-label emotion recognition, allowing for the identification of one or more emotional states within each open-ended response. To achieve this, we employ a RoBERTa model fine-tuned on the *GoEmotions* dataset (further implementation details are discussed in Section 5.2). While classic categorical models typically rely on Ekman's six *basic* emotions, GoEmotions aligns with the componential view by offering a high-granularity taxonomy of 27 distinct categories and a neutral one, making 28 in total. Specifically, the taxonomy comprises: a*dmiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness,* and *surprise*, alongside the *neutral* category. This granularity is essential for analyzing precariousness, as it allows for the disentangling of complex negative states—distinguishing, for instance, between *anger* at the system and *fear* of the future.

2.3.2 Dimensional emotion model

While the categorical approach identifies specific emotional labels, this study also employs a Dimensional Approach to capture the underlying affective properties of the text within a continuous vector space.

We utilize the *VAD model*, a foundational framework that posits any emotional state can be described by three numerical dimensions:

- Valence (V): Measures the degree of pleasure, ranging from negative (1) to positive (5). In the context of precariousness, this dimension helps quantify the severity of the negative sentiment expressed by the respondent. High valence corresponds to states like joy or satisfaction, while low valence characterizes sadness, anger, or disgust.
- Arousal (A): Measures the physiological and psychological intensity, from calm (1) to excited (5). This distinction is critical for differentiating between active dissatisfaction (high arousal) and resignation or hopelessness (low arousal).
- Dominance (D): Measures the perceived level of control over a situation, from submissive (1) to dominant (5). High dominance is associated with feelings of empowerment, pride, or being in control, while low dominance characterizes states of being powerless, guided, or overwhelmed by external forces.

To implement this, we use a BERT-based regression model trained on the *EmoBank* corpus. Unlike the classification task performed by RoBERTa, this model assigns a precise 3-dimensional coordinate to each response, providing a nuanced geometric representation of the respondents' affective states (further information of the implementation in Section 5.2).

3. Data description and pre-processing

3.1. The VulnYouth dataset

The dataset used in this study—*VulnYouth: Studying Intersectional Vulnerability of Precarious Youth in Spain* (see Maestripieri et al., 2025)—is publicly available through the *Portal de la Recerca de Catalunya*[1]. It consists of *3012 interviews* conducted in February 2023 with Spanish youth aged *20 to 34*. The sample is a non-probabilistic quota sample, stratified by gender, age, migration background and urban/rural residence. To ensure adequate representation of the most vulnerable young people—often underrepresented in online surveys—210 face-to-face interviews were conducted in disadvantaged neighbourhoods in Catalonia. The remaining interviews were collected through an online self-administered questionnaire, which took respondents approximately 16 minutes to complete.

The questionnaire was designed to be as exhaustive as possible in capturing the multiple operationalizations of precariousness. It contains sociodemographic indicators (for example, gender, age, migration background), the economic and employment status (for example, working time, type of contract, and income), the level of education and several metrics of mental health, such as the *WHO-5 Well-Being Index*. This index is a widely used screening tool for depression and subjective well-being.

A fundamental component of the survey and the focus of this thesis is an open question that asks respondents to provide their own definition using free text: *"What does precariousness mean to you?"* This question, which had no text limits, was designed to capture the subjective and personal understanding of the phenomenon, going beyond the predefined indicators. A total of *2487 valid, non-empty answers* to this open question were obtained from the total 3012 interviews. This corpus of 2487 short text responses constitutes the main data for our analysis. While previous studies have tried to understand which factors best represented the WHO-5 mental health index (see Venturini, 2024) and perceived precariousness (see Di Liberatore, 2024), this project will focus on detecting the emotions of open responses of the survey. The objective is to move beyond *what* young people associate with precariousness and to analyze *how* they feel about it.

3.2. Key variables: the open-ended question

The crucial variable that we focus on in our study of the VulnYouth dataset is $V54\_PRECARIOUSNESS$, which captures open answers to the question *"What is precariousness in your opinion?"*. Thanks to the work of prof. Lara Maestripieri and Elena Di Liberatore (see Di Liberatore, 2024), the answers—originally written in Spanish—were fully translated into English, allowing the use of machine learning models in the English language, and were stored in a unique file (*english_document.xlsx*) containing the valid responses.While the dataset provides a substantial sample size of 2487 valid responses, a major challenge is represented by the short length of the text data. The responses average 19.25 English words per response (standard deviation 22.06). The distribution of these word counts is illustrated in Figure 1. This brevity poses a significant challenge for automated

[1]The full dataset and documentation are available at: `https://doi.org/10.34810/data1821`

analysis. As highlighted by Nandwani and Verma (2021), short and unstructured texts often lack the necessary context for machines to effectively comprehend human psychology, leading to ambiguities where individuals fail to clearly express specific emotional states (see Nandwani and Verma, 2021).

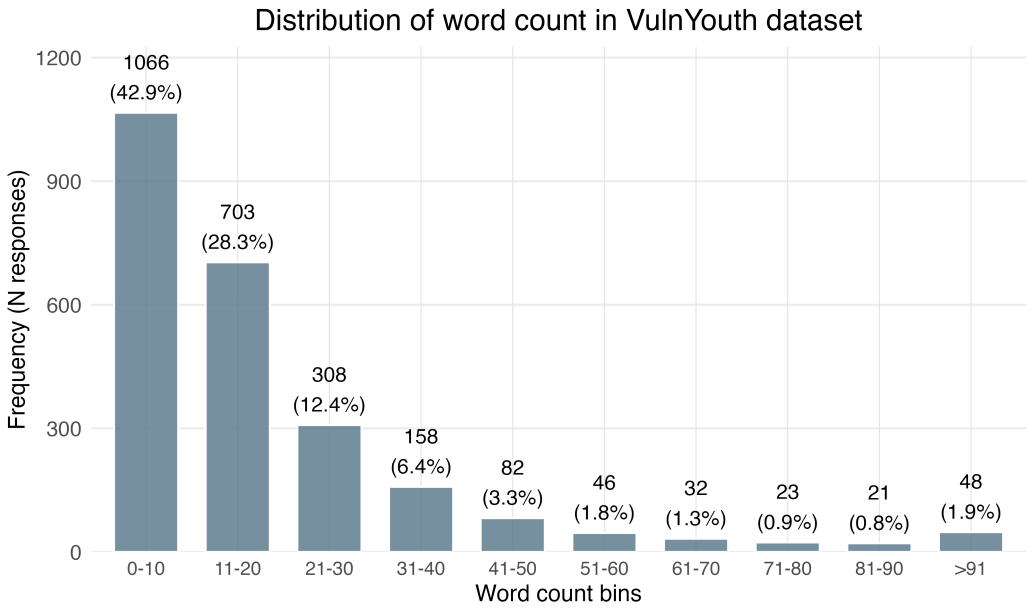

Figure 1: Distribution of word count in $V54\_PRECARIOUSNESS$ (VulnYouth). Data refers to the $N = 2487$ valid answers, translated from Spanish to English.

This issue is compounded by the specific nature of the topic. Unlike general emotion detection, which often deals with datasets balanced between positive and negative examples, the subject of precariousness yields a corpus almost exclusively focused on *negative experiences*. Therefore, the analytical task is not simply to classify sentiment, but to make fine-grained distinctions between a variety of negative emotional states—such as anger, fear, sadness, and disappointment—from limited textual information. The existing literature offers few established frameworks for this specific combination of challenges. This study, therefore, aims to contribute by presenting a *multi-method approach* as a case study for analyzing nuanced emotions in short, affectively skewed, open-ended survey responses (see Abdullah Al Maruf et al., 2024).

3.3. Text pre-processing pipeline

Two distinct pre-processing pipelines were implemented in this study, necessitating the use of two different programming environments. R was utilized for the Structural Topic Model (STM) to ensure full reproducibility with the previous work by Di Liberatore (2024) and to leverage the native `stm` package, which is the standard implementation for this methodology. Conversely, Python was selected for the BERT-based classifiers to exploit the advanced deep learning libraries (such as PyTorch or TensorFlow) and the GPU acceleration capabilities offered by platforms like Google Colab and Kaggle, which are essential for training Large

Language Models.

3.3.1 STM Pre-processing

The first key preliminary step was data alignment. Starting from the raw translated responses and the full VulnYouth sociodemographic dataset, it was necessary to ensure that each text answer matched the correct respondent's metadata. This alignment process, built upon the data cleaning framework established by Di Liberatore (2024), involved merging the text documents from *english_documents.xlsx* with the respondent metadata using a unique ID key, ensuring no mismatches occurred due to missing responses.

At this point, the aligned text corpus was processed using the `textProcessor` function of R's `stm` package. The pipeline involved the following steps:

1. we preserve specific punctuation, notably hyphens (e.g., `custompunctuation = c("-")`), which allowed the model to treat complex concepts as single units rather than splitting them into separate, less informative words.
2. Stop word removal: Common English stop words (e.g., *and*, *the*) were removed thanks to a customized list defined in the previous thesis. Additionally, the specific terms *precariousness* and *precarious* were removed as custom stop words to prevent them from dominating every topic and obscuring more specific themes.
3. Stemming: Words were reduced to their root form (e.g., *instability* becomes *instabil*).
4. Removal of punctuation and numbers: All punctuation and numeric characters were removed.
5. Lowercasing: All text was converted to lowercase.

This process resulted in a processed object (stored in *stm_preprocessed_input.RData*) containing two essential components for the model: the vocabulary, which is the list of all unique valid words remaining after the cleaning process, and the documents list, containing the indices and associated frequencies of these words for each response.

3.3.2 BERT pre-processing

Unlike the STM approach, pre-trained language models like BERT rely on the full linguistic structure of sentences to infer meaning. Therefore, for the BERT and derived RoBERTa-based analysis, we adopted a less destructive pre-processing strategy.

Specifically, we did not perform stemming or stop-word removal, as these elements provide essential contextual cues for the self-attention mechanisms of the Transformer architecture.

The text cleaning pipeline, applied consistently to both the training datasets (GoEmotions and EmoBank) and our target corpus (*english_documents.xlsx*), included the following steps to ensure compatibility with the `bert-base-uncased` tokenizer:

- Lowercasing: All text was converted to lowercase.
- Noise removal: HTML tags, web URLs, and non-essential characters such as emojis and special symbols were stripped.
- Whitespace normalization: Essential punctuation was preserved, while multiple whitespaces were consolidated into single spaces.

Finally, the external datasets were formatted for the fine-tuning phase—the process of adapting a general-purpose pre-trained model to a specific downstream task (discussed in detail in Section 3.3.2). For this purpose, the 28 emotion columns from GoEmotions were consolidated for the categorical classification task, while the Valence, Arousal, and Dominance columns from EmoBank were used as continuous targets for the dimensional regression task.

4. Methodology

This chapter details the technical implementation of the three different methods used in this study, starting from the Structural Topic Model.

4.1. Structural Topic Modeling

The Structural Topic Model (STM) is a probabilistic generative model of word counts that extends Latent Dirichlet Allocation (LDA) by incorporating document-level covariates into the prior distributions of topic prevalence and topical content. Unlike heuristic approaches, STM allows for the formal estimation of the relationship between document metadata (covariates) and latent topic structures.

Figure 2 provides a graphical representation of the model, distinguishing between the generative process and the estimation via variational inference.

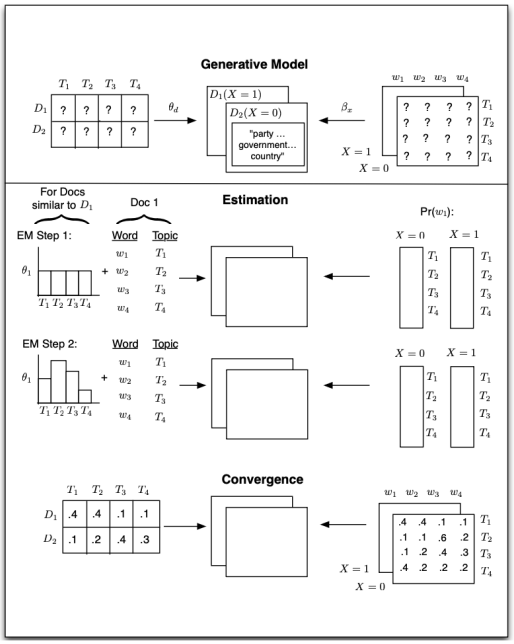

Figure 2: Heuristic description of generative process and estimation of the STM. Source: Roberts et al. (2019)

4.1.1 Generative process

The generative process assumes a corpus of D documents with a vocabulary of size V and a fixed number of topics K. The model defines two core components influenced by parameters

Topic Prevalence (θ_d) and *Topical Content* (β_k).

Topic prevalence: The proportion of a document d allocated to each topic is denoted by θ_d, which is a vector of length K residing on the simplex (i.e., $\sum_{k=1}^{K} \theta_{d,k} = 1$). This vector is drawn from a Logistic Normal distribution conditional on the document covariates X_d:

$$\theta_d | X_d\gamma, \Sigma \sim \text{LogisticNormal}(\mu = X_d\gamma, \Sigma) \tag{1}$$

Formally, this distribution is defined by drawing a latent variable η_d from a multivariate normal distribution in R^{K-1} and mapping it to the simplex via the softmax function:

$$\eta_d \sim \mathcal{N}^{K-1}(X_d\gamma, \Sigma), \qquad \theta_{d,k} = \frac{\exp(\eta_{d,k})}{\sum_{j=1}^{K} \exp(\eta_{d,j})} \tag{2}$$

where $\eta_{d,K}$ is fixed to 0 for identifiability. The components are defined as follows:

- X_d is a $1 \times P$ vector of document-level covariates (including an intercept).
- γ is a $P \times (K-1)$ matrix of regression coefficients, mapping the covariates to the topic simplex.
- Σ is a $(K-1) \times (K-1)$ covariance matrix capturing the correlations between topics.

Topical content: The distribution of words within topic k, denoted as $\beta_{d,k}$, represents the probabilities of the Multinomial distribution over the vocabulary V from which the words are drawn. Unlike standard models, these probabilities can vary by covariates. Modeled via a distributed multinomial regression framework, the generative probability for a word v in document d and topic k is:

$$\beta_{d,k,v} \propto \exp(m_v + \kappa^{(t)}_{k,v} + \kappa^{(c)}_{y_d,v} + \kappa^{(i)}_{y_d,k,v}) \tag{3}$$

where parameters represent log-space deviations from a baseline word frequency m_v: $\kappa^{(t)}$ is the topic-specific deviation, $\kappa^{(c)}$ is the covariate-group deviation, and $\kappa^{(i)}$ is the interaction effect.

The generative process for the words is defined as follows: for each position n in document d, a latent topic indicator $z_{d,n}$ is first drawn from the document's topic proportions θ_d. Conditional on this topic assignment (i.e., given $z_{d,n} = k$), the observed word $w_{d,n}$ is drawn from a Categorical (or Multinomial with size 1) distribution over the vocabulary:

$$w_{d,n} | z_{d,n}, \beta_{d,k} \sim \text{Categorical}(\beta_{d,k=z_{d,n}}) \tag{4}$$

Priors and regularization. To ensure identification and prevent overfitting, the model imposes regularizing prior distributions on the global parameters. In this study, we employ the default prior specifications of the `stm` software, which are designed to balance model complexity with data fit.

For *Topic Prevalence prior* (γ), a "Pooled" prior is applied. This assumes a hierarchical structure defined as $\gamma \sim \mathcal{N}(0, \sigma^2)$ with $\sigma \sim$ Half-Cauchy$(1, 1)$. This configuration induces moderate shrinkage, effectively assuming that the effect of the covariate is small unless the

data strongly indicates otherwise.

For *Topical Content*, the regularization applies to the collection of deviation coefficients, denoted generally as κ (encompassing $\kappa^{(t)}$, $\kappa^{(c)}$, and $\kappa^{(i)}$). An L1 Lasso prior is used:

$$\kappa \sim \text{Laplace}(0, 1/\lambda) \tag{5}$$

This prior promotes sparsity by forcing the coefficients of irrelevant word-covariate interactions to exactly zero, ensuring that only significant vocabulary deviations are retained in the model structure. The regularization parameter λ is automatically selected by the `glmnet` algorithm included in `stm`, which utilizes the AIC information criterion to optimize the trade-off between data fit and model sparsity. This prior forces the coefficients of irrelevant words to exactly zero; consequently, only the vocabulary that significantly varies across covariate groups retains non-zero coefficients.

4.1.2 Estimation: variational E-M

Since the exact posterior inference is intractable, we approximate Variational Expectation-Maximization (EM) algorithm.

1. Initialization: Global parameters (γ, Σ, κ) are initialized, typically using spectral decomposition for stability.
2. E-Step (Variational Inference): In this step, the global parameters are treated as fixed. The algorithm estimates the approximate posterior distribution for the document-level latent variables: the topic proportions η_d (where $\theta_d = \text{softmax}(\eta_d)$) and the word-topic assignments $z_{d,n}$.
3. M-Step (Maximization): In this step, the algorithm updates the point estimates for the global parameters (γ, Σ, κ, β) to maximize the regularized lower bound on the marginal likelihood (ELBO). This effectively performs Maximum a Posteriori (MAP) estimation given the sufficient statistics collected in the E-step.
4. Convergence: The process repeats until the relative change in the variational objective drops below a specified tolerance.

The final result of this algorithm, which constitutes the basis for our analysis, consists of:

- The MAP point estimates for the relationship between covariates and topics (†) and the topic correlations ($\hat{\Sigma}$).
- The Expected Values of the topic proportions for every document ($(\hat{\theta})$.).
- The MAP point estimates for the word usage within topics ($(\hat{\beta})$.), derived from the estimated ($\hat{\kappa}$). coefficients.

4.2. BERT

BERT (Devlin et al., 2019) is a language model based on the Transformer architecture, designed to generate contextual representations of text. It is commonly referred to as a Large Language Model because it is associated with large-scale neural networks. It has become popular due to its ease of application in various fields, and this flexibility is achieved through so-called transfer learning: instead of starting from scratch, which would be expensive and require too much data, it uses parameters obtained during a pre-training phase, where the

model learns the English language from massive corpora. An output layer is then added to adapt the new model to specialize in a specific task, such as in our case predicting the three continuous variables Valence, Arousal, and Dominance for EmoBank, or predicting the 28 different categorical variables for GoEmotions.

4.2.1 The architecture: Encoder

Large Language Models are divided by functional roles into encoder, decoder, and encoder-decoder. Although architectures used in the past included *Long Short-Term Memory* (LSTMs) or *Convolutional Neural Networks* (CNNs)—models designed to process sequences sequentially or extract local features—the current state of the art, including BERT, is based on the *Transformer* architecture (see Vaswani et al. 2017). Specifically, BERT utilizes the Encoder component of the Transformer. This choice is due to the Self-Attention mechanism's ability to overcome the parallelization and long-term memory limitations of previous models. While Decoder architectures are oriented towards generation (NLG) and Encoder-Decoder architectures towards translation, Encoders are designed to project the input, i.e., the text, into a latent space that preserves deep semantics. Such latent representations materialize, for each token, into output vectors $v \in R^D$ (e.g., $D = 768$ for BERT-Base). Here, represents the hidden size dimension, so named because it refers to the size of the internal neural layers where the model constructs abstract representations that are "hidden" from the raw input and output layers. Defined as *Contextualized Word Embeddings* these vectors are not fixed—unlike traditional Static Embeddings—but vary dynamically based on syntactic and semantic relationships with other words in the sentence. We will now provide an overview of how a Transformer is structured.

4.2.2 The Transformer Encoder

The structure of the Encoder is defined by three main sequential components, as described in Figure 3:

1. Input Embedding: The process begins with the raw input text—which, in our study, corresponds to the open-ended survey responses regarding precariousness. The input is transformed into vectors that combine the token's meaning, its position in the sentence, and the segment it belongs to. The sum of these three elements constitutes the input for the first block (details in the next section).
2. Stacked Transformer Blocks: This is the core of the architecture. It consists of identical blocks in sequence (from 12 to 96 in current LLMs). Each block follows the original Transformer design proposed by Vaswani et al.(2017), a multilayer network composed of two main sub-layers: a Multi-Head Self-Attention mechanism (which contextualizes the information) and a Position-wise Feed-Forward Network (FFN). Each sub-layer is followed by a residual connection and a Layer Normalization according to the formula:

$$\text{Output} = \text{LayerNorm}(x + \text{Sublayer}(x)) \tag{6}$$

   The final output is a sequence of highly contextualized vectors (Contextualized Embeddings).
3. The Output Head (Task Specific): The output of the last block is processed differently

depending on the phase:

- During Pre-training: BERT uses two heads in parallel:
  (a) Masked LM Head: Projects the masked token vectors into the vocabulary space (via Softmax) to recover the original token.
  (b) NSP Head: Uses the vector corresponding to the special token for a binary classification, determining if the second sentence logically follows the first (Next Sentence Prediction).
- During Fine-tuning: The pre-training heads are removed and replaced with a specific Classification Head (e.g., a linear layer followed by Softmax) usually applied on top of the token for tasks such as whole-sentence classification.

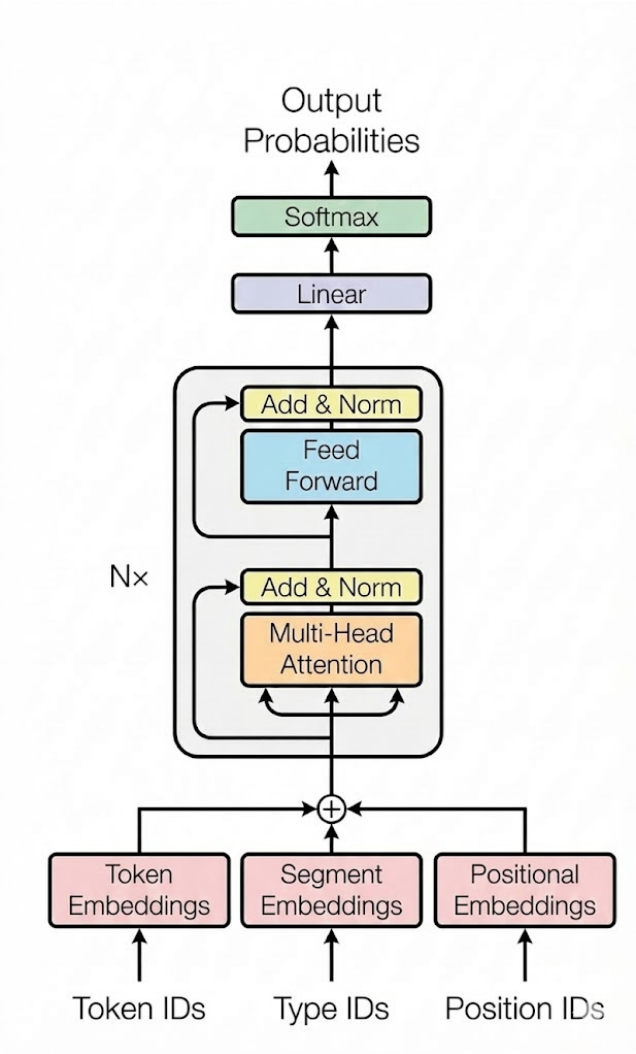

Figure 3: BERT architecture

4.2.3 Input embedding

The BERT input representation is the sum of three distinct components: *Token Embeddings*, *Segment Embeddings*, and *Position Embeddings*.

Let B denote the batch size and L be the sequence length (the fixed length of the input

sequence). The first step is Tokenization. BERT uses the *WordPiece* algorithm, which is based on a *Greedy Longest-Match-First* strategy: given a word, the tokenizer looks for the longest subunit present in the pre-trained reference vocabulary V. If the word does not exist, it is decomposed into subunits (e.g., "*playing*" $\rightarrow$ "*play*","*##ing*", where the "##" prefix indicates a non-leading subword); if no subunit is valid, the special token $[UNK]$ is assigned. The vocabulary V consists of 30522 tokens originally generated from the concatenation of the *Toronto BookCorpus* and *English Wikipedia* (see Devlin et al., 2019).

Subsequently, input formatting takes place to obtain three index matrices of dimension $[B, L]$:

1. Token IDs: The token sequence is enriched with the special tokens $[CLS]$ (start of sequence) and $[SEP]$ (separator between sentences).
2. Segment IDs (or Type IDs): This is a matrix of zeros and ones. 0 is assigned to all tokens in the first sentence of the token sequence (including $[CLS]$ and the first $[SEP]$), while 1 is assigned to the tokens of the following sentence.
3. Position IDs: This is an incremental counter (0, 1, ..., $L-1$) indicating the token's position in the sequence.

To ensure that all inputs have length L, the Truncation and Padding process is applied:

- If the sequence has length $N > L$, it is truncated.
- If $N < L$, padding tokens (ID 0) are added until L is reached.

These indices are not processed directly by the network but pass through an Embedding Lookup. Each index i is used to retrieve the corresponding dense vector from a weight matrix $W \in R^{V \times D}$, where V is the vocabulary size and D is the hidden size.

Mathematically, this operation is equivalent to the product between a one-hot vector $x_{one-hot} \in \{0,1\}^V$ (where the element at index i is 1 and the others are 0) and the matrix W:

$$e = x_{one-hot} \cdot W$$

However, computationally, this multiplication is inefficient. In real implementations (e.g., PyTorch or TensorFlow), an *indexing* or *slicing* operation is performed, directly extracting the i-th row of the matrix:

$$e = W[i, :]$$

The final embedding for each token t is given by the element-wise sum of the three components, followed by layer normalization to stabilize training. This step applies the transformation $x' = \frac{x-\mu}{\sigma} \cdot \gamma + \beta$.

$$E^{(t)}_{input} = \text{LayerNorm}(E^{(t)}_{token} + E^{(t)}_{segment} + E^{(t)}_{pos})$$

Consequently, the complete input representation fed into the first Encoder layer is a tensor of dimension $[B, L, D]$ (Batch Size $\times$ Sequence Length $\times$ Hidden Size).

4.2.4 Stacked Transformer blocks

As shown in Figure 4, the Stacked Transformer Block is composed of two main sub-layers: a Multi-Head Attention mechanism and a Feed-Forward Network (FFN).

In modern architectures, a Pre-Layer Norm configuration is often used: normalization is applied to the input before entering the sub-component, while the residual connection adds the original input directly to the component's output. The operation for each sub-layer can be described as:

$$x_{out} = x + \text{Sublayer}(\text{LayerNorm}(x)) \tag{7}$$

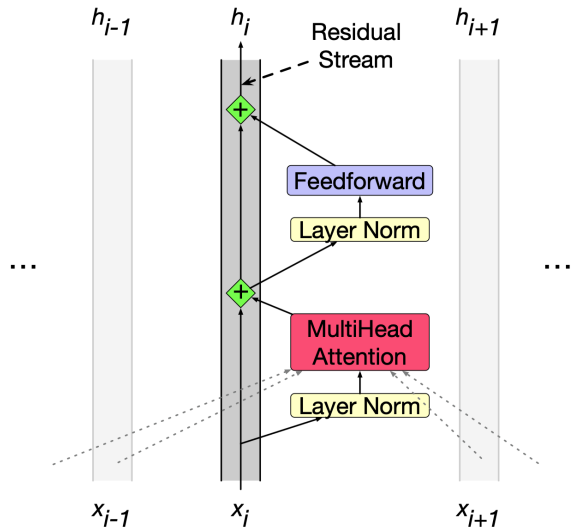

Figure 4: Detail of the Residual Stream and the Pre-Layer Norm configuration within the block for the BERT algorithm. Source: Jurafsky and Martin (2024).

4.2.5 Multi-head attention

The input of this module is the tensor H (coming from the embeddings or the previous block) of dimension $[B, L, D]$, where D is the hidden size (or d_{model}). The mechanism allows the model to jointly focus on different parts of the input from different representation subspaces. Instead of computing a single attention function over the entire dimension D, the input tensor is linearly projected h times in parallel (e.g., $h = 8$). Each projection reduces the dimension to $d_k = D/h$ (e.g., $768/8 = 96$), generating h independent "heads". The h heads work in parallel. Their outputs (each of dimension $[B, L, d_k]$) are finally concatenated to restore the original dimension D and linearly projected via the matrix W^O to obtain the final result.

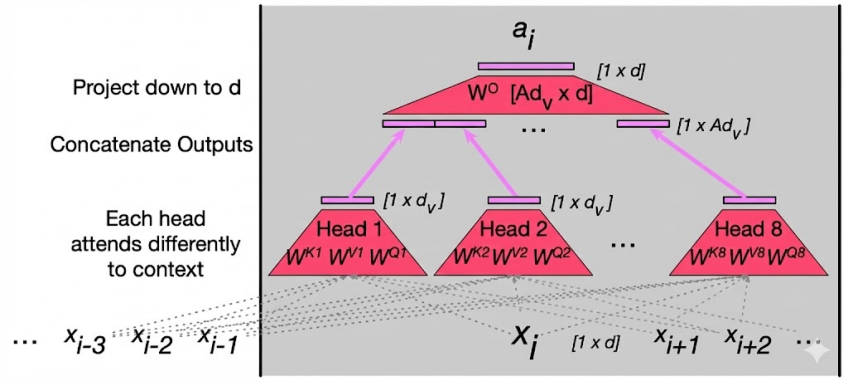

Figure 5: Diagram of Multi-Head Attention in Encoder configuration (bidirectional context) for the BERT algorithm. Source: Jurafsky and Martin (2024).

Self-Attention (scaled dot-product) Inside each head, the attention calculation takes place. The trainable weight matrices $W^Q, W^K, W^V \in R^{D \times d_k}$ transform the input tensor into three new distinct tensors of dimension $[B, L, d_k]$:

- Query (Q): Represents the current token seeking information.
- Key (K): Represents the token against which the comparison is made.
- Value (V): Contains the actual information that will be extracted.

To understand the mechanism at the single-token level, let us consider the row vectors x_i and x_j extracted from the tensor. The projection vectors are calculated as:

$$q_i = x_i W^Q, \quad k_j = x_j W^K, \quad v_j = x_j W^V \tag{8}$$

The attention score, which determines how much "importance" to give to token j when processing i, is calculated via the dot product between Query and Key $score(x_i, x_j) = q_i \cdot k_j^T$. It is subsequently scaled by the square root of the key dimension ($\sqrt{d_k}$) for numerical stability, and finally normalized via Softmax:

$$\alpha_{i,j} = \text{softmax}\left(\frac{score(x_i, x_j)}{\sqrt{d_k}}\right) \tag{9}$$

The output of the head for token i is the weighted sum of the Value vectors:

$$\text{Head}_i = \sum_j \alpha_{i,j} v_j \tag{10}$$

Finally, the heads are concatenated and projected to return to the dimension $[B, L, D]$:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h) W^O \tag{11}$$

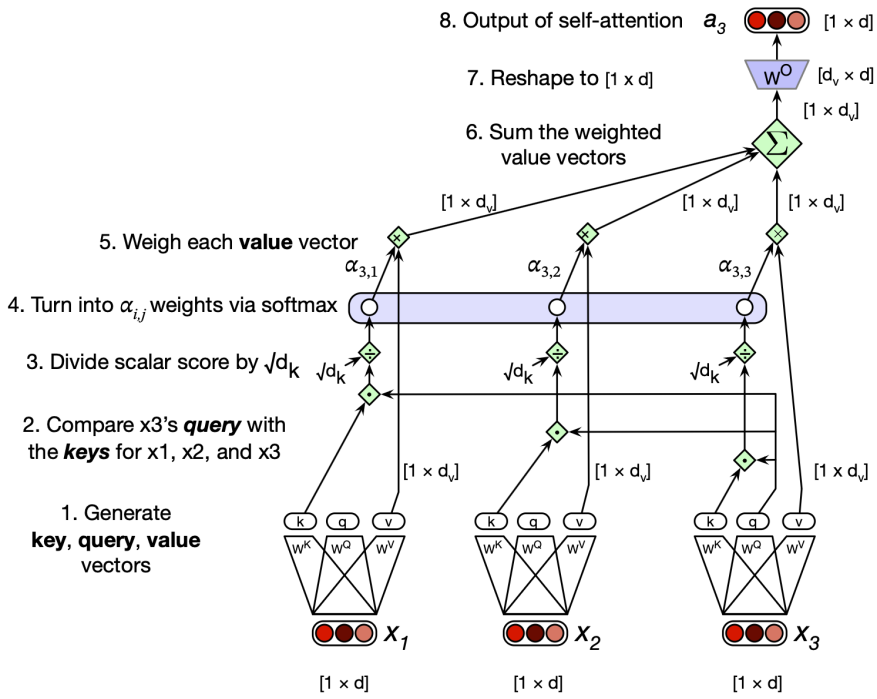

Figure 6: Calculating the value of a3, the third element of a sequence using self-attention. Source: Jurafsky and Martin (2024).

4.2.6 Feed-Forward network

A Feed-Forward Neural Network (FFNN), commonly known in the literature as a Multilayer Perceptron (MLP) (Goodfellow et al., 2016), is composed of multiple layers of interconnected neurons that apply affine transformations followed by non-linear activation functions. In the following example, we see a simple 2-layer feedforward network with one hidden layer, one output layer, and one input layer. The formulas are: $h = \sigma(Wx + b)$
$z = Uh$
$y = softmax(z)$
Where:

- x is the input vector, of dimension n_0
- W is the weight matrix from the input layer to the hidden layer, of dimension $[n_0, n_1]$
- b is the bias vector of the hidden layer, of dimension $[n_1]$
- σ is the non-linear activation function (e.g., sigmoid or ReLU $= \max(0, x)$)
- h is the vector of hidden layer activations, of dimension $[n_1]$
- U is the weight matrix from the hidden layer to the output layer, of dimension $[n_2, n_1]$
- z, called the logit, is the linear output of the final level, of dimension $[n_2]$
- y is the final output (e.g., class probabilities via softmax), of dimension $[n_2]$

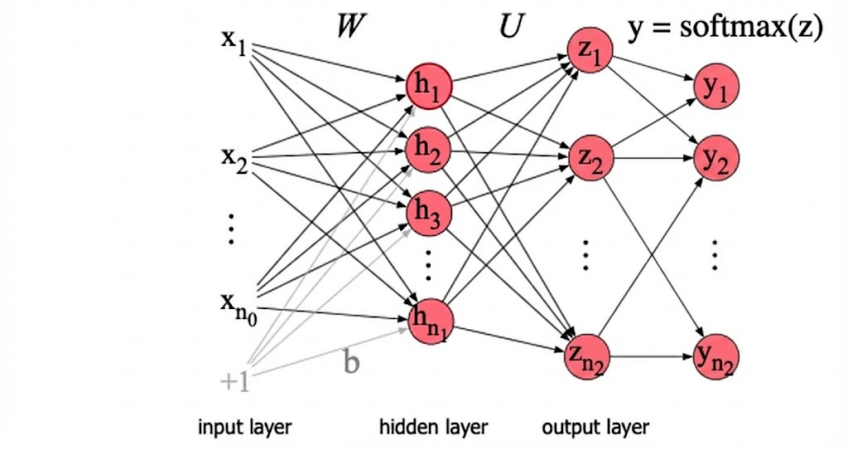

Figure 7: A simple 2-layer feedforward network

Since FFNN is a supervised machine learning method, we know the correct output y for every observation x. During training, we want to minimize the output error. To find the right E, W, and U—that is, the parameters we will use for inference—we use error propagation, which relies on gradient descent optimization.

In BERT, the feedforward network takes the specific name of Position-wise Feed-Forward Network. The crucial difference compared to the generic example is that this network does not process the entire sentence in one go to classify it, but is applied identically and separately to every single position (token) of the sequence, acting on the vectors exiting the previous level, the self-attention. If we call x the vector of a single input token to the FFN, the operation is described by the formula:

$$FFN(x) = GELU(xW_1 + b_1)W_2 + b_2$$

.Where:

- The input x has dimension d_{model} (e.g., 768).
- The first linear transformation expands the vector into a higher-dimensional space ($d_{ff} = 4 \times d_{model}$, e.g., 3072) via W_1 and b_1.
- The GELU (Gaussian Error Linear Unit) activation function is applied, which performs better than ReLU ($= \max(0, x)$) on deep transformer models.
- The second linear transformation projects the vector back to the original dimension d_{model} via W_2 and b_2.

The output of the FFN is not the final result (y), but is summed back to the input x (residual connection) and normalized (Layer Normalization) before passing to the next Encoder block. In summary, while Attention "looks" at other words to build context, the FFN "processes" that context for each word individually.

During pre-training In this phase, BERT is trained using multi-task learning: for each batch of data entering the output heads, the model simultaneously calculates both loss

functions. The total Loss that is minimized is the sum of the two Losses:

$$Loss_{total} = Loss_{MLM} + Loss_{NSP}$$

Thanks to these two processes, BERT is able to learn bidirectional contextual representations.

- **Masked Language Modeling (MLM) Head:** Standard language models predict the next word in a sequence, limiting context to one direction (left-to-right). To enable bidirectional learning, BERT masks a percentage of the input tokens and attempts to predict them based on the surrounding context.
  The process follows a specific stochastic procedure described in the literature (see Jurafsky et al., 2024):
  1. Selection: The model randomly selects 15% of the token positions in the input sequence.
  2. Masking Strategy: Among the selected tokens, the model does not always replace them with the $[MASK]$ token. Instead, it applies the following rule to prevent a mismatch between pre-training (where masks are present) and fine-tuning (where they are not):
     - 80% of the time: The token is replaced with the $[MASK]$ token (e.g., *"the man went to the store"* $\rightarrow$ *"the man went to the [MASK]"*).
     - 10% of the time: The token is replaced with a random word (e.g., $\rightarrow$ *"the man went to the penguin"*). This forces the model to keep track of the context to detect inconsistencies.
     - 10% of the time: The token remains unchanged (e.g., $\rightarrow$ *"the man went to the store"*). This biases the representation towards the actual word.

  The output of the MLM head is a dense layer projecting to the vocabulary size V, followed by a Softmax function. The model then minimizes the Cross-Entropy Loss between the resulting probability distribution and the one-hot encoding of the original token.
- **Next Sentence Prediction (NSP) Head:** Many downstream tasks, such as Question Answering (QA) or Natural Language Inference (NLI), require understanding the relationship between two sentences. To learn this, BERT is fed pairs of sentences (A, B) during training.
  - 50% of the time: B is the actual sentence that follows A in the corpus (labeled as *IsNext*).
  - 50% of the time: B is a random sentence from the corpus (labeled as *NotNext*).

  The vector corresponding to the $[CLS]$ token passes through a binary classification layer followed by Softmax to predict the probability that B follows A.

At the end of this phase, the *outcome* is a set of pre-trained parameters (weights) that capture deep syntactic and semantic features of the language, ready to be fine-tuned for specific tasks.

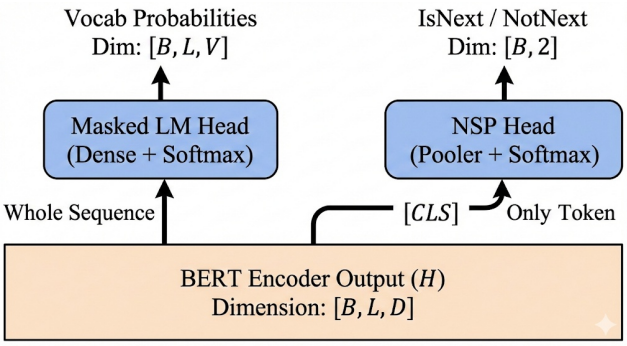

Figure 8: Output head during pre-training

During fine-tuning For downstream tasks like Emotion Detection, the pre-training heads are discarded and replaced by a task-specific head. Since our objective is sentence-level classification (or regression for EmoBank), the model ignores the individual token embeddings $h_1, ...h_L$ along the sequence length dimension $[L]$ and utilizes only the aggregate representation provided by the $[CLS]$ token at index 0.

Consequently, the input to the final Output Head is a matrix of dimension $[B, D]$. This vector is processed by a new linear layer:

- For Multi-label Classification (e.g., GoEmotions): The layer projects to dimension $[B, K]$ (where $K = 28$ is the number of classes), followed by Sigmoid. Here each sentence is not restricted to a single unique category but can be classified into multiple emotions simultaneously (e.g., expressing both anger and sadness). The output is a probability distribution across all 28 categories.
- For Regression (e.g., EmoBank): The layer projects to dimension $[B, 3]$ (Valence, Arousal, Dominance) with a suitable activation function.

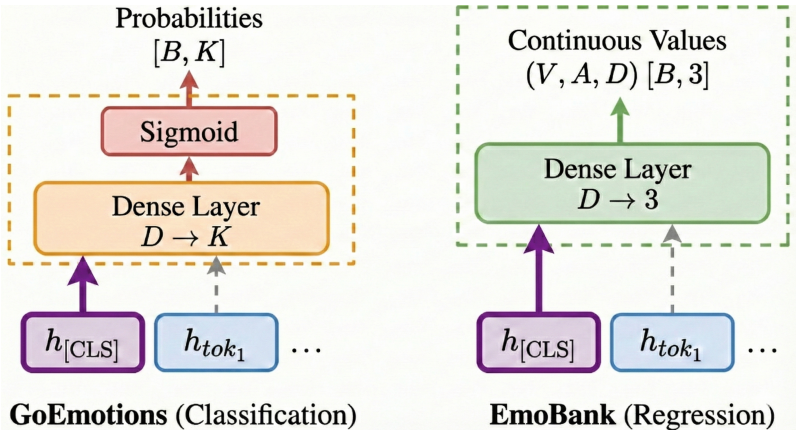

Figure 9: Fine tuning for GoEmotions on the left and Emobank on the right

Final inference output Once trained, these two distinct models are applied to our target dataset of open-ended survey responses (VulnYouth), yielding two different types of outputs for each respondent:

1. The categorical model returns a vector of 28 probabilities, indicating the likelihood

of each specific emotion being present in the response (allowing us to detect mixed emotional states).
2. The dimensional model returns 3 continuous numerical values representing the estimated Valence, Arousal, and Dominance levels of the entire response.

4.3. Models used: RoBERTa and BERT

Although BERT represents the reference standard in the literature, this work adopts two specific architectural variants. The selection of these models was not arbitrary but the result of a preliminary comparative analysis conducted during the experimental phase. We tested different Transformer-based architectures to identify the best performer for each distinct task:

- RoBERTa (Robustly optimized BERT approach) for Categorical Classification: For the GoEmotions dataset, we empirically observed that RoBERTa [Liu et al., 2019] outperformed the standard BERT model. In our validation tests, it achieved superior metrics in terms of Accuracy and F1-Score, likely benefiting from its larger vocabulary and dynamic masking during pre-training.
- BERT (Bidirectional Encoder Representations from Transformers) for Dimensional Regression: Conversely, for the EmoBank dataset (VAD regression), the standard BERT architecture provided the most accurate estimates for the continuous variables (Valence, Arousal, Dominance), minimizing the Mean Squared Error compared to other variants.

4.3.1 Technical comparison

The structural and training differences between the base model (BERT) and the adopted variants are summarized in the table below.

RoBERTa distinguishes itself by the elimination of the *Next Sentence Prediction* task (deemed superfluous), the use of *Dynamic Masking* (the mask changes at every epoch), and training on a data corpus ten times larger with much larger Batch Sizes.

Feature	BERT-Base	RoBERTa-Base
Pre-training Data Size	16 GB	160 GB
Vocabulary (V)	30,522	50,265
Model Dimension (d)	768	768
Transformer Layers	12	12
Attention Heads	12	12
Max Sequence Length (L)	512	512
Batch Size (B)	256	8,000
Total Parameters	$\approx$ 110M	$\approx$ 125M

5. Model configuration

After having defined the theoretical frameworks of Structural Topic Modeling and Transformer-based architectures, this section details the specific experimental setup adopted to apply these methods to the VulnYouth dataset. We first address the selection process for the Structural Topic Model, focusing on the quantitative metrics used to determine the optimal number of latent topics (K) that best balances semantic coherence and exclusivity, and the choice to introduce different covariates in the topic prevalence field. Subsequently, we outline the fine-tuning strategy for the BERT and RoBERTa models, detailing the training hyper-parameters, the data splitting procedures for validation, and the specific metrics employed to evaluate performance in both the categorical (GoEmotions) and dimensional (EmoBank) emotion analysis tasks.

5.1. STM model configuration

The topic modeling analysis was conducted on the corpus of open-ended responses regarding the perception of precariousness. The final dataset for this specific method consists of 2472 documents (valid responses) with a vocabulary size of V= 861 unique terms. To identify the optimal latent structure, we estimated two distinct Structural Topic Models:

1. Base Model: This model replicates the exact configuration adopted by Di Liberatore in her previous analysis of the dataset. It is an unsupervised model with no covariates, serving as a baseline. We retained this model to ensure continuity with the semantic labeling and topic definitions previously established by Di Liberatore and Prof. Maestripieri, allowing for a consistent comparison of results (see Di Liberatore, 2024).
2. Covariate Model: A model incorporating document-level metadata to estimate topic prevalence. Specifically, we included five binary covariates reflecting the respondents' socio-economic conditions:
    - *dum_ecoind*: Economic independence from the family of origin.
    - *dum_jobinsec*: Perception of job insecurity.
    - *dum_endmeet*: Difficulty in making ends meet.
    - *dum_normepres*: Perception of social pressure.
    - *dum_unexpmix*: Ability to handle unexpected expenses.

    We adopted binary variables to isolate specific risk profiles identified as significant in the literature, grounding our operationalization in the framework of Maestripieri (2024). For *dum_ecoind*, derived from a 5-point Likert scale regarding future independence, we isolated the highest category to identify respondents who perceive themselves as fully economically independent from their parents. Regarding *dum_jobinsec*, the original variable measured the probability of losing one's job on a 4-level scale; we recoded the "Likely" and "Very Likely" responses as 1 to capture high-intensity insecurity. For *dum_endmeet*, which originally assessed the difficulty of making ends meet on a 6-level scale, we selected the two most severe categories ("Great difficulty" and "Difficulty") to define acute economic strain. *dum_unexpmix* was derived from binary indicators of financial capacity, coded as 1 for those with the individual ability to pay unexpected expenses and 0 for those dependent on others. Finally, *dum_normepres* was based on the Employment Precariousness Scale (EPRES), normalized to a range

of 0 to 1; we applied a strict threshold, coding only values exceeding 0.8 as vulnerable to isolate cases of acute perceived precarity.

These covariates were employed exclusively to model *topic prevalence*. The *topical content* was kept independent of covariates, relying on a *global content* assumption. This decision was driven by statistical parsimony. Given the moderate sample size (D) relative to the number of latent topics ($K = 20$), introducing content covariates would have required estimating separate vocabulary distributions for each covariate group, creating 20 topics for two different sets of data. Such fragmentation would likely lead to data sparsity issues, resulting in unstable word-probability estimates (β). Therefore, adhering to a global content assumption allows for more robust inference by pooling information across all respondents to define the semantic meaning of the topics.

The number of topics was fixed at $K = 20$. This decision was primarily driven by the need to ensure methodological consistency with the baseline model established by Di Liberatore. By retaining the same parameter K, we allow for a direct comparison between Di Liberatore's results and the current analysis.

In the previous analysis, Di Liberatore identified $K = 20$ as the optimal solution after performing a diagnostic search across a range of candidates ($K \in [14, 25]$). While statistical fit metrics such as held-out likelihood and residuals suggested multiple viable candidates, the final decision was made to maximize Semantic Coherence: this metric measures the tendency of the most probable words in a topic to co-occur within the same documents. At $K = 20$, the model exhibited the highest coherence, ensuring that the resulting topics were the most semantically consistent and interpretable for qualitative analysis.

To understand the qualities balanced in this selection, we recall the definitions of the two primary diagnostic metrics:

- **Semantic coherence**: This metric measures the tendency of the most probable words in a topic to co-occur within the same documents. High coherence ensures that the topic is semantically consistent and human-interpretable. Formally, for a set of the M most probable words in topic k, coherence is defined as (Mimno et al., 2011):

$$C(k) = \sum_{m=2}^{M} \sum_{l=1}^{m-1} \log \frac{D(v_m^{(k)}, v_l^{(k)}) + 1}{D(v_l^{(k)})} \tag{12}$$

  where $D(v)$ is the document frequency of word v, and $D(v, v')$ is the co-document frequency of words v and v'.
- **Exclusivity**: This metric quantifies the extent to which the top words for a topic are confined to that topic and unlikely to appear in others. Exclusivity is formalized via the FREX metric, which balances the word's frequency within the topic and its exclusivity relative to other topics:

$$\text{Exclusivity}(k) \propto \sum_{v \in V} \left(\frac{\omega}{\text{ECDF}(\beta_{k,v})} + \frac{1 - \omega}{\text{ECDF}(\beta_{k,v} / \sum_j \beta_{j,v})} \right)^{-1} \tag{13}$$

  where $\beta_{k,v}$ is the probability of word v in topic k, and ECDF denotes the Empirical Cumulative Distribution Function.

5.2. BERT fine-tuning strategy for emotion detection

To adapt the pre-trained Transformer models to the specific tasks of categorical and dimensional emotion analysis, we employed a fine-tuning approach using the *HuggingFace Transformers* library. While the underlying transfer learning mechanism remains consistent, the training configurations, loss functions, and evaluation metrics were tailored to the distinct nature of the *GoEmotions* (multi-label classification) and *EmoBank* (regression) datasets.

Categorical emotion analysis

For the categorical emotion classification, we apply a RoBERTa-base model. The selection of RoBERTa over BERT is justified by the empirical results obtained during the training phase, where it achieved higher validation scores compared to the latter. This superior performance is attributed to RoBERTa's optimized pre-training procedure, which has been shown to handle complex multi-label dependencies more effectively.

Training setup We utilized the standard *GoEmotions* data splits (Train/Val/Test) to ensure comparability with the literature. The model was trained for 8 epochs with a batch size of 32, using the AdamW optimizer and a learning rate of 2e-5. A crucial design choice was the implementation of the *MultiLabel Focal Loss*. Preliminary experiments using the standard *BCEWithLogitsLoss* resulted in poor inference performance on our target dataset, *VolunYouth*, where over 80% of the samples were classified as "neutral." The Focal Loss addresses this by dynamically down-weighting easy examples and forcing the model to focus on hard-to-classify instances. This shift prevented the model from collapsing into the majority class, significantly reducing the neutral bias.

Validation and optimization strategy Following the training phase, we implemented a Dynamic Thresholding strategy to optimize the classification performance. In multi-label tasks, the standard fixed probability threshold (e.g., 0.5) is often suboptimal for rare classes. Therefore, we evaluated the model on the *validation set* to calculate the F1-score across a range of possible probability values. We then identified and selected the specific probability cut-off that maximized the F1-score for each of the 28 emotions independently. These optimized thresholds were fixed and subsequently applied to the test set.

Test set results Given the multi-label nature of the task and significant class imbalance, we relied on the F1-Score as the primary performance indicator, computing four distinct variations:

- Macro F1: this metric calculates F1-score independently for each emotion classes and then takes the unweighted mathematical average.
- Micro F1: This metric aggregates the total number of true positives, false negatives, and false positives globally across all classes, and then calculates a single F1-score. It gives more weight to the majority classes.
- Weighted F1-Score: This is the weighted version of the Macro F1 score, accounting for class imbalance.

- Samples F1-Score: This metric is specific to multi-label classification. Instead of looking at performance column-by-column per emotion, it calculates the F1-score for each individual data instance row-by-row and then averages them.

As shown in Table 1, our RoBERTa model achieved a Macro F1-score of 0.5742. This result is highly satisfactory, significantly outperforming the original BERT-base baseline (Macro F1: 0.46) reported by (Demszky et al., 2020) confirming the effectiveness of our Focal Loss and dynamic thresholding strategy.

Metric	**Score**
F1 Micro	0.6271
F1 Macro	0.5742
F1 Weighted	0.6268
F1 Samples	0.6314

Table 1: Performance metrics on the GoEmotions test set (28 labels).

Inference on VulnYouth Once validated, the optimized RoBERTa model was deployed to perform inference on the 2487 responses of the VulnYouth corpus. By applying the emotional thresholds identified during the validation phase, we generated a binary multi-label matrix where each respondent's definition of precariousness is associated with one or more of the 28 emotional categories. This allows for the detection of complex, mixed emotional states, such as the simultaneous presence of sadness and disappointment.

Dimensional emotion analysis (BERT on EmoBank)

For the continuous mapping of emotions, we fine-tuned a BERT-base-uncased model to perform a multi-output regression task, predicting the Valence, Arousal, and Dominance (VAD) scores simultaneously.

Training setup We utilized the standard *EmoBank* data splits to ensure comparability with the literature (see Buechel & Hahn (2017)). The model was trained for 8 epochs with a batch size of 32 and a learning rate of 2e-5. A critical adaptation in this phase was the choice of the loss function. Instead of the standard Mean Squared Error (MSE), we implemented the *Concordance Correlation Coefficient (CCC) Loss*, described by this formula:

$$Loss = 1 - \text{CCC} = 1 - \frac{2\rho\sigma_{\hat{y}}\sigma_y}{\sigma_{\hat{y}}^2 + \sigma_y^2 + (\mu_{\hat{y}} - \mu_y)^2}$$

This metric penalizes predictions that deviate from the original observations in both variance and mean.

Validation and model selection To prevent overfitting and select the optimal regression weights, at the end of each training epoch we evaluated the model on the *EmoBank validation set*. We monitored the average CCC score across the three dimensions (V, A, D) and saved the model of the epoch that achieved the highest global validation CCC.

Test set results Finally, we evaluated the optimized model on the unseen *test set* using three complementary metrics:

- Mean Absolute Error (MAE): To quantify the average absolute distance between prediction and truth.
- Pearson Correlation (r): To measure the linear relationship.
- Concordance Correlation Coefficient (CCC): To measure the agreement (accuracy + precision).

The results, detailed in Table 2, demonstrate a significant reduction in Mean Absolute Error (MAE), achieving approximately half the error magnitude reported for human inter-annotator agreement by Buechel & Hahn (2017). The "human-level agreement" (IAA) measures how much one human annotator correlates with the average of the others. Regarding the Pearson correlation (r), our model aligns with state-of-the-art findings: it achieves a higher result for Valence (0.79 vs. 0.74) while yielding slightly lower results for Arousal and Dominance. This discrepancy is consistent with the literature, as these dimensions are notoriously difficult to infer from text alone without the support of acoustic or visual cues (Buechel & Hahn, 2017).

Metric	Average	Valence	Arousal	Dominance
Our Model MAE	0.1816	0.1797	0.1966	0.1686
Human Agreement MAE	0.3860	0.3490	0.4410	0.3670
Our Model Pearson (r)	0.6091	0.7872	0.5429	0.4973
Human Agreement (r)	0.6340	0.7380	0.5950	0.5700
Our Model CCC	0.6042	0.7779	0.5394	0.4953

Table 2: Regression performance on the EmoBank test set. Human Agreement values (Pearson and MAE) are derived from the 'Reader' perspective in Buechel & Hahn (2017).

Inference on VulnYouth The fine-tuned BERT regression model was subsequently applied to the target dataset to map each open-ended response into the VAD continuous space. For every respondent, the model estimated three numerical coordinates representing Valence, Arousal, and Dominance. These scores provide a geometric representation of the affective experience, enabling a quantitative comparison of emotional intensity and perceived agency across different socio-demographic groups.

6. Results

This section describes the inference from applying the methods described in Section 4 to the *VulnYouth* open-ended questions. The results are presented in four parts: first, we provide a descriptive analysis of the textual corpus; second, we examine the underlying emotional dimensions (Valence, Arousal, Dominance) using the BERT regression model; third, we analyze the specific emotions using the categorical RoBERTa model; finally, we explore the thematic content using Structural Topic Modeling (STM);

6.1. Descriptive analysis of the corpus

Before applying emotional models, we provide a descriptive analysis of the lexical frequencies to understand the vocabulary young people use to define precariousness. The corpus contains 2487 valid documents with a total of 22902 tokens; the vocabulary has 1,738 distinct words.

The textual data underwent a rigorous cleaning process as detailed in Section 3.3.1. During this stage, we applied a customized pipeline using the `textProcessor` function from the `stm` package.

The most frequent stems in the corpus are *work* (1083 occurrences), *abl* (734) (being able or ability), *live* (637), *job* (501), *salari* (480) (salary or salaries) , and *condit* (461)(condition). The dominance of *work*, *job*, and *conditions* signals that the labour market is the primary aspects of precariousness. Moreover, the high frequency of *able* - stemmed in *abl* - and *live* points out that precarity is defined by the inability to sustain a life independent of family support.

Responses are generally brief, with an average of 9 tokens per document after processing. This brevity leads to a highly sparse *Document-Term Matrix.* A *Document-Term Matrix* is a mathematical structure representing a corpus of text. In this matrix, rows correspond to individual documents and columns correspond to unique terms (vocabulary). Each cell $X_{i,j}$ contains a value—typically a frequency count or a weighted score—indicating the occurrence of term j within document i.To quantify the lexical overlap between these individual accounts, we calculated the mean cosine similarity.

The cosine similarity between two document vectors, A and B, is defined as:

$$\text{similarity}(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

- A_i and B_i: Represent the frequency of term i in documents A and B.
- The numerator $(A \cdot B)$: Is the dot product, representing the common terms shared by both responses.
- The denominator $(\|A\|\|B\|)$: Is the product of the Euclidean norms, which normalizes the score to a range between 0 and 1.

The average similarity across all pairs in the corpus is extremely low (0.07). This result is statistically significant for our methodological choice: it proves that respondents use a highly heterogeneous vocabulary to describe their conditions. Simple word-matching or frequency counts are insufficient because young people rarely use the exact same terms. This justifies the use of the Structural Topic Model (STM), which does not rely on direct lexical overlap but uncovers latent thematic patterns by analyzing how words co-occur across different documents.

6.1.1 Distinctive terms (TF-IDF) and mental health

To look beyond simple frequency and identify terms that characterize specific experiences, we applied *Term Frequency–Inverse Document Frequency (TF-IDF) weighting*: this metrics allow us to identify terms that are less frequent globally but highly distinctive to specific

responses. Formally, the weight $w_{t,d}$ for a term t in a document d is defined as the product of its local frequency and its global rarity:

$$w_{t,d} = tf_{t,d} \times idf_t = tf_{t,d} \times \log\left(\frac{N}{df_t}\right) \tag{14}$$

where:

- $tf_{t,d}$ is the raw count of term t in document d.
- N is the total number of documents in the corpus.
- df_t is the number of documents containing term t.

Distinctive terms include compounds like *salary-to-expense*, *nutrition*, anxd *four-walls*, which evoke the struggle to manage basic needs. Moreover, terms like *depress*, *distress*, *anxiety*, and *uncertainty* appear as distinctive markers, that using simple frequency counts had missed, and they reveal that precarity is experienced as a chronic strain on psychological well-being and biographical stability. We observe that, while the statistical analysis was conducted on stemmed roots to reduce lexical sparsity, the results are reported here in their full linguistic forms to ensure clarity and interpretability.

6.2. Dimensional emotion analysis

The findings from the previous descriptive stage—specifically the high lexical heterogeneity (low cosine similarity) and the brevity of the responses—highlight the limitations of purely frequentist approaches. While respondents use a wide variety of specific terms to describe their lives, these linguistic fragments likely point toward a shared underlying affective state. To capture this deeper commonality that simple word-counts miss, the following section moves from lexical analysis to a dimensional emotional analysis. By mapping these sparse documents into a continuous emotional space (see Section 2.3.2), we can determine if the observed linguistic diversity masks a more uniform psychological experience of precariousness.

6.2.1 General distribution of affect

The global distribution of scores confirms the negative and disempowering nature of the discourse on precariousness.

- Valence: The mean valence score is 2.56 ($SD \approx 0.35$). As shown in Figure 10, the density plot is skewed to the left, quantitatively confirming that precariousness is conceptualized almost exclusively through negative sentiment.
- Arousal: The mean arousal score is 2.95 ($SD \approx 0.32$), placing the discourse very close to the neutral threshold. The low standard deviation suggests a concentration of responses around this neutral point; this lack of emotional intensity is consistent with the categorical findings, which showed a negligible presence of high-arousal emotions such as anger (0.5%) or fear (1.6%)..
- Dominance: the mean dominance score is 2.73 ($SD \approx 0.28$). This distribution is shifted towards the lower end of the scale of Dominance, indicating that respondents perceive precariousness as a state of low agency, controlled by external forces rather

than their own choices (see Mohammad (2018)).

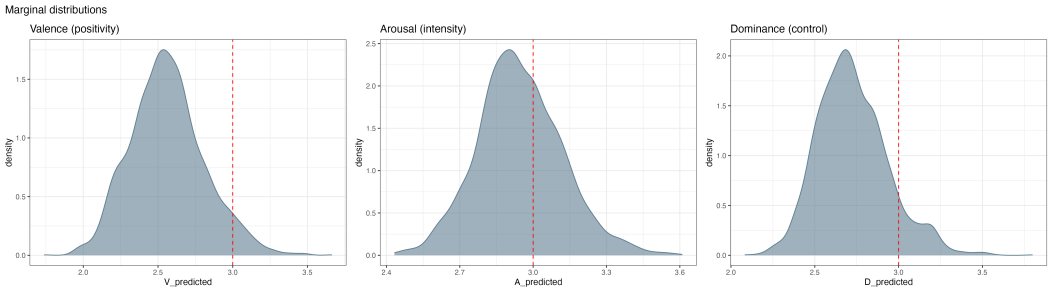

Figure 10: Marginal distributions of emotional dimensions. The red dashed line indicates the neutral score of 3.

To complement the marginal distributions, the pairwise relationships between the emotional dimensions are illustrated in the following scatterplots (Figure 11).

- Valence vs. Arousal: The plot reveals a dense cluster in the lower-left quadrant, representing low-valence and low-to-moderate arousal states. This concentration confirms that the negative sentiment associated with precariousness is not typically expressed through high-intensity agitation, but rather through a more muted, resigned affect, consistent with the low prevalence of active emotions like anger or fear.
- Valence vs. Dominance: A strong positive correlation is visible, where lower valence scores consistently align with lower dominance scores. This indicates that the more negative the respondent's definition of precariousness, the more it is characterized by a perceived lack of control and agency, grounding the sociological definition of precarity as a state of powerlessness.
- Arousal vs. Dominance: The distribution is centered around the intersection of the neutral thresholds (score of 3), with a slight bias toward the lower-dominance region. The absence of data points in the high-arousal/high-dominance quadrant further underscores the lack of empowerment or active conflict in the narratives, reinforcing the "neutrality paradox" observed in the categorical analysis.

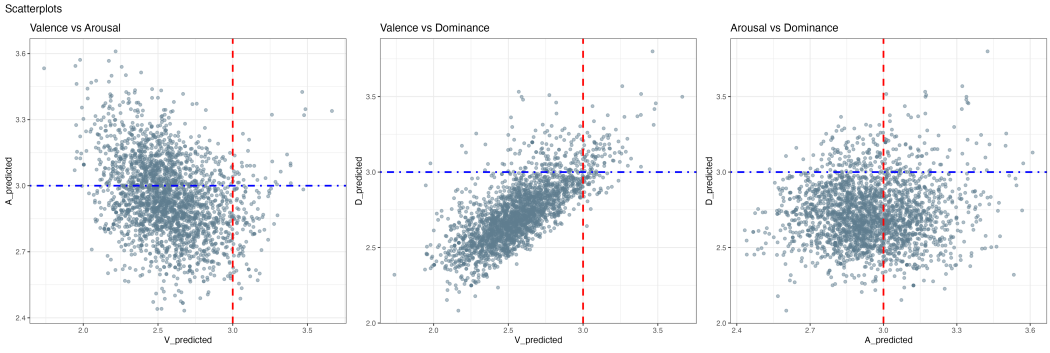

Figure 11: Pairwise scatterplots of the predicted emotional dimensions (Valence, Arousal, and Dominance) for the VulnYouth corpus. The red and blue dashed lines represent the neutral states

6.3. Categorical emotion analysis

Whe categorical analysis was conducted by applying a RoBERTa model fine-tuned on the GoEmotions dataset (see Section 5.2). This approach identifies specific emotional labels associated with precariousness, classifying each response into one or more of 28 distinct categories.(see Section 2.3.1).

6.3.1 Prevalence of emotional categories

Figure 12 presents the frequency of the 28 detected emotions in the VulnYouth open-ended question. The distribution reveals an emotional landscape dominated by feelings of resignation rather than agitation.

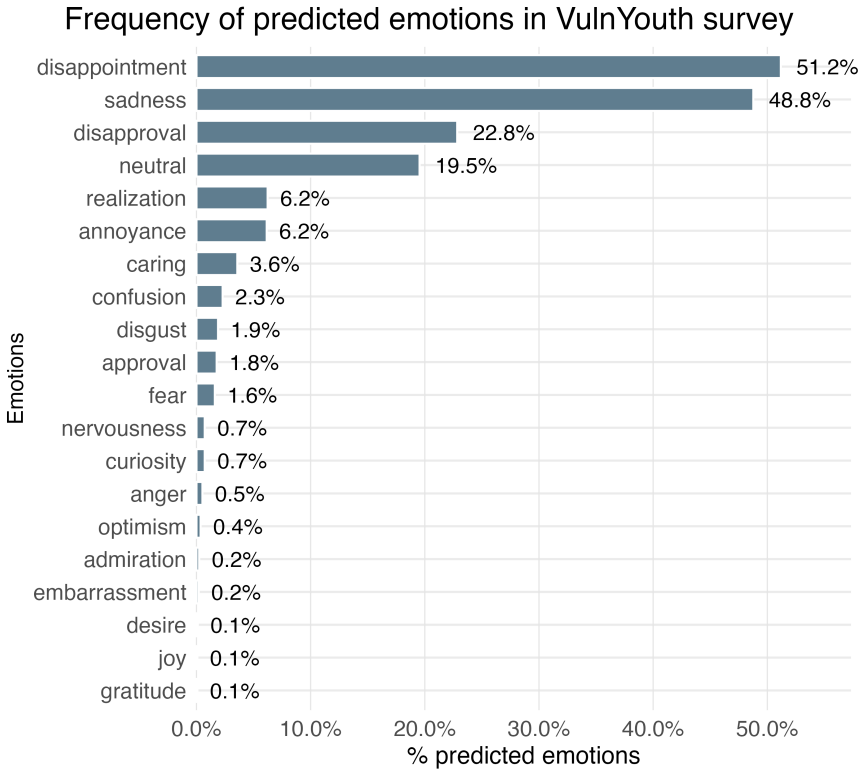

Figure 12: Frequency of predicted emotions in the open-ended questions of precariousness in the VulnYouth survey.

Based on the data presented in Figure 12, the emotional landscape of the corpus is defined predominantly by disappointment and sadness. The prevalence of these states, alongside the notable absence of high-arousal emotions, suggests a shared narrative of resignation rather than active conflict.

- *Disappointment* (51.2%): This is the most prevalent emotion, appearing in over half of the responses. Within the GoEmotions taxonomy, disappointment is distinguished from generic sadness as "sadness or displeasure caused by the nonfulfillment of one's hopes or expectations". This implies a cognitive, comparative judgment between a desired future and a failed reality, rather than a purely passive state.

- *Sadness* (48.8%): Closely following disappointment, sadness captures the generic emotional pain or sorrow expressed by respondents. The high correlation between these two states suggests they often merge into a complex state of "sad resignation".
- *Disapproval* (22.8%): A significant portion of respondents express an "unfavorable opinion," indicating that many young people are actively judging the current socio-economic system as inadequate or wrong.
- Absence of high-arousal Negativity: Consistent with the dimensional VAD findings, high-arousal negative emotions are remarkably rare. Active states such as Anger (0.5%) and Fear (1.6%) appear with negligible frequency. This aligns with the "low dominance" finding: the prevailing mood is not one of (*anger*) or panic (*fear*), but of resignation (*sadness*) and unmet expectations (*disappointment*).

6.3.2 Emotional co-occurrence and complex states

The multi-label nature of the RoBERTa model allows us to observe how emotions overlap within a single response. To investigate the relationships between these states, Figure 13 presents a Pearson correlation heatmap calculated on the binary presence vectors of the emotional labels. The analysis focuses specifically on the top five most frequent emotions (Disappointment, Sadness, Disapproval, Neutral, and Realization). While the GoEmotions taxonomy includes 28 distinct categories, a full 28×28 correlation matrix would suffer from high sparsity and limited interpretability.

The heatmap in Figure 13 reveals some structural patterns in how respondents articulate precariousness:

- There is a strong positive correlation (indicated by the deep red block) between *Disappointment* and *Sadness*. This confirms that these two states rarely appear in isolation; rather, respondents articulate a complex state of sad resignation, where the feeling of loss (Sadness) is intrinsically tied to a failed expectation (Disappointment).
- Conversely, the *Neutral* category shows a strong negative correlation with both Disappointment and Sadness. This can indicate a binary linguistic behavior: respondents either adopt a descriptive, detached tone (*Neutral*) or fully commit to the emotional narrative of resignation.
- Ultimately, the categories *disapproval* and *realisation* do not show a significant correlation with other emotions.

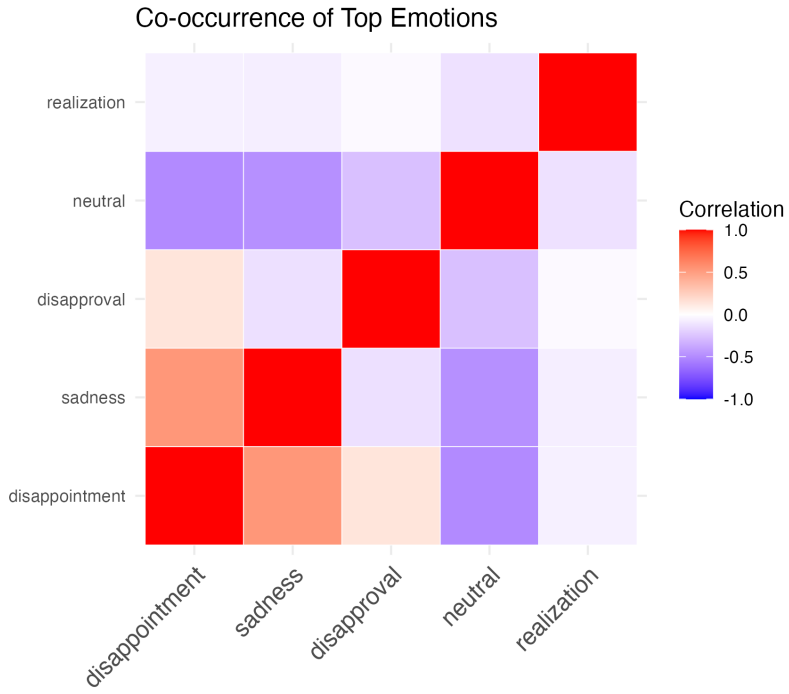

Figure 13: Co-occurrence heatmap of the top 5 predicted emotions. The colors represent Pearson correlation coefficients calculated from binary detection vectors; red indicates a positive association, while blue indicates a negative association.

6.3.3 Dividing the 28 emotions in 4 sentimental groups

To synthesize the complex emotional landscape, the 28 categorical emotional labels were aggregated into four high-level sentiment groups based on their affective valence and composition:

- Positive group: composed of 15 emotions associated with favorable or constructive states (*admiration, amusement, approval, caring, curiosity, desire, excitement, gratitude, joy, love, optimism, pride, realization, relief, surprise*).
- Negative group: composed of 12 emotions associated with unpleasant states, distress or aversion (*anger, annoyance, confusion, disappointment, disapproval, disgust, embarrassment, fear, grief, nervousness, remorse, sadness*).
- Ambivalent group: containing responses that simultaneously exhibit at least one positive and one negative emotion, reflecting a complex or mixed emotional state.
- Neutral group: containing responses classified exclusively as "neutral" or lacking any specific emotional marker.

Figure 14 illustrates the distribution of these groups across the corpus. The Negative group clearly dominates the affective narratives identified in the open-ended responses, confirming that the subjective conceptualization of precariousness is overwhelmingly rooted in distress.

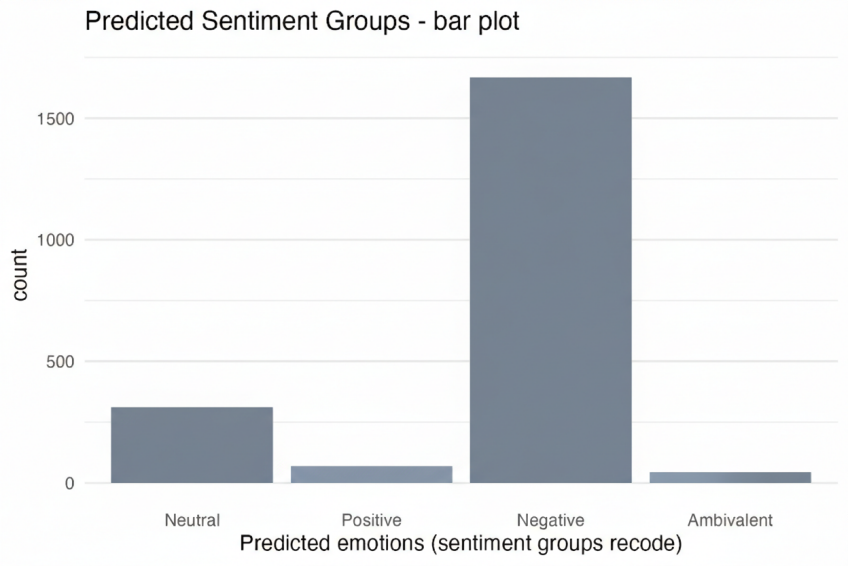

Figure 14: Distribution of aggregated sentiment groups. Note the dominance of negative sentiment and the significant minority of neutral responses.

6.3.4 The neutrality paradox: a logistic regression analysis

To formally test whether socio-demographic and material vulnerability factors are associated with emotional expression, we estimated two binary logistic regression models: respondents were coded as emotional (emotions_dum = 1 assigned to Positive, Negative, or Ambivalent sentiment groups) versus neutral (emotions_dum = 0 exclusively Neutral category). The logit model estimates the log-odds of being classified as emotional.

Formally:

$$\text{logit}\big(P(Y=1)\big) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k \tag{15}$$

where $Y = 1$ indicates emotional expression and $X_1, \ldots, X_k$ represent the set of predictor variables included in the model (e.g., gender, age group, migration background in Model 1; and employment vulnerability indicators in Model 2). The coefficients β_k capture the change in the log-odds of being classified as emotional associated with a one-unit change in each predictor.

Model 1: Socio-demographic predictors The first model includes the standardized variables (gender, age group, and migration background) as categorical predictors, and their transformations are detailed in Appendix A:

$$\text{emotions\_dum} \sim \text{gender} + \text{age\_dum} + \text{migr\_dum} \tag{16}$$

Gender emerges as a statistically significant predictor of emotional expression ($\beta = 0.305$, $p = 0.0039$). Exponentiating the coefficient: $\exp(0.305) \approx 1.36$ This indicates that women have approximately 36% higher odds of expressing emotional language compared to men

(reference category), holding age and migration background constant.

Age and migration background are not statistically significant ($p > 0.3$ and $p > 0.4$, respectively), indicating no systematic variation across these dimensions.

Model 2: Objective vulnerability predictors and the neutrality paradox
The second model examined whether material conditions predict emotional expression:

$$\begin{aligned} \text{emotions\_dum} \sim \text{dum\_normepres} + \text{dum\_jobinsec} + \text{dum\_endmeet} \\ + \text{dum\_unexpmix} + \text{dum\_ecoind} \end{aligned} \tag{17}$$

These variables represent objective vulnerability indicators: normative employment status, job insecurity, difficulty making ends meet, unexpected financial strain, and economic independence. They follow the variable standardization established in Section 5.1 and completely outlined in Appendix A.

The model shows that objective job insecurity is the only statistically significant predictor ($\beta = -0.285$, $p = 0.0207$). Exponentiating: $\exp(-0.285) \approx 0.75$

Respondents experiencing high job insecurity have approximately 25% lower odds of being classified as emotional compared to those with stable employment conditions, while the other vulnerability indicators are not statistically significant ($p > 0.10$).

Contrary to expectations, higher levels of objective job insecurity are associated with a lower probability of emotional expression. Those in more unstable labour conditions are more likely to produce linguistically neutral definitions of precariousness.

We refer to this pattern as the *neutrality paradox*: individuals exposed to higher structural instability are less likely to frame their condition using explicit emotional language.

The neutrality paradox emerges as a statistically significant association within the present dataset. Whether this pattern reflects emotional normalization, strategic linguistic framing, or broader socio-cultural mechanisms remains an open empirical question that warrants further investigation.

6.4. Topic analysis: thematic structure and emotional drivers

Having established the emotional landscape of the corpus, we apply Structural Topic Modeling (STM) to map the thematic content of the responses. We estimated two distinct models to identify the latent themes of the open-ended responses:

- Baseline model: A standard unsupervised model without covariates, replicating the configuration used in previous research by Di Liberatore (2024) to establish a semantic starting point.
- Covariate model: An extended model that incorporates document-level metadata—specifically *Employment Vulnerability, Job Insecurity, Economic Strain, and Economic Independence*—as predictors of topic prevalence. This allowed us to estimate how these specific life conditions influence the probability of a respondent discussing certain themes.

Semantic stability To ensure methodological continuity, we compared the results of our Covariate Model with the Baseline Model. As shown in Table 3, the most influential semantic definitions of the topics remain stable. Table 3 specifically compares the top three highest-probability words for a subset of topics between the two models. The high degree of overlap (e.g., Topic 5: "hour", "contract", "long" in both models) confirms that the addition of covariates does not shift the core meaning of the topics, providing a robust foundation for further analysis.

Topic	Base Word 1	Our Word 1	Base Word 2	Our Word 2	Base Word 3	Our Word 3
3	work	work	poor	condit	environ	poor
5	hour	hour	contract	contract	long	long
8	hous	minimum	minimum	hous	food	access
16	basic	basic	salari	salari	money	expens
18	salari	low	low	salari	wage	wage

Table 3: Comparison of Top Words between the Baseline Model (Di Liberatore, 2024) and the Covariate Model (Current Study). The semantic overlap confirms model stability.

Model quality and fit We further evaluated the models using two diagnostic metrics:

- Held-Out Likelihood: We calculate the log-likelihood of the models on a held-out portion (10%) of the data. The Covariate Model achieved a likelihood of -5.150, statistically equivalent to the Baseline Model's -5.149. This indicates that the added complexity of the covariates does not degrade the model's generalizability.
- Coherence vs Exclusivity: Figure 15 illustrates the trade-off between Semantic Coherence (word co-occurrence) and Exclusivity (distinctiveness). The Covariate Model (Red) occupies the same semantic space as the Baseline Model (Blue), with some topics (e.g., T16, T9) even achieving higher coherence. This confirms that accounting for emotional framing maintains, and in some cases clarifies, the thematic structure.

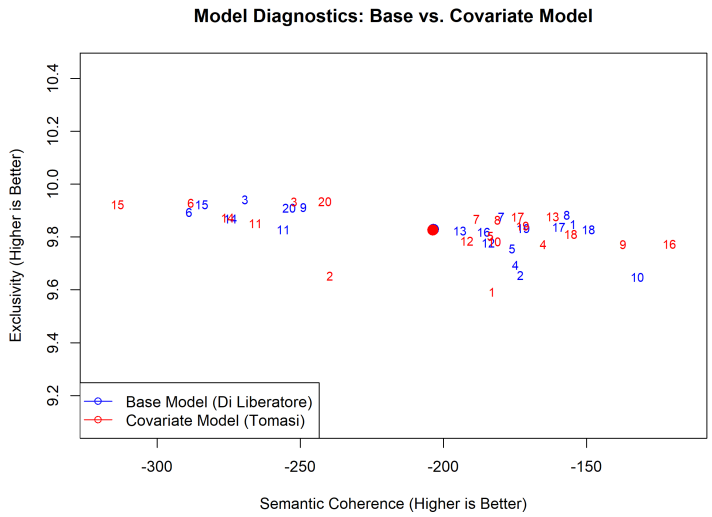

Figure 15: Diagnostic comparison: Semantic Coherence vs. Exclusivity. The overlap indicates that adding covariates preserves topic quality.

6.4.1 The 20 latent themes of precariousness

Table 4 presents the full list of 20 topics identified in the corpus. The labels used here are validated labels from previous research, meaning they retain the exact thematic definitions established by Di Liberatore and Maestripieri (see Di Liberatore, 2024) to allow for direct cross-study comparisons. The "Key Terms" in this table are derived from our Covariate Model, which, as previously defined, is the model incorporating socio-economic indicators to weight the prevalence of each theme.

ID	Label (Di Liberatore, 2024)	Key Terms (Highest Prob & FREX)
T1	Resource-driven economic insecurity	*famili, take, studi, care*
T2	Unaffordable independent living	*pay, dont, rent, place*
T3	Poor working conditions	*work, poor, condit, environ*
T4	Worker exploitation and vulnerability	*surviv, support, difficult, precari*
T5	Overtime without proper compensation	*hour, contract, long, temporari*
T6	Illegal and unsafe working conditions	*condit, live, dignifi, money*
T7	Not being valued in the workplace	*paid, person, day, remuner*
T8	No access to housing or food	*hous, minimum, food, access*
T9	Inability to meet living standards	*live, save, suffici, digniti*
T10	Impossibility to save money	*save, famili, buy, afford*
T11	Living an unstable life	*earn, good, know, lead*
T12	Impossibility to plan a future	*life, expens, stabil, secur*
T13	Struggle to make ends meet	*meet, end, time, make*
T14	Lack of resources	*lack, resourc, opportun, subsist*
T15	Exploitation and Abusive Treatment	*bad, pay, qualiti, treatment*
T16	Not being able to afford basic needs	*basic, salari, money, expens*
T17	Impossibility of reaching economic independence	*econom, independ, situat, young*
T18	Low wages	*salari, low, wage, pay*
T19	Youth unemployment and insecurity	*time, econom, peopl, stabl*
T20	Job insecurity and insufficient stability	*job, kind, standard, mistreat*

Table 4: The 20 latent topics of precariousness in the VulnYouth corpus. Semantic labels are adopted from Di Liberatore (2024) to ensure cross-study interpretability, while the reported key terms (Highest Probability and FREX) are derived from the current covariate model that incorporate the 5 socio-economic indicators

Emotional predictors of topics Beyond identifying the 20 latent themes, we examine whether specific predicted emotions are systematically associated with the probability of discussing particular topics. The analysis focuses on the four most frequent emotions—sadness, disappointment, disapproval, and neutrality—alongside fear.

The methodology follows a two-stage approach. First, we established a ranking of the most prevalent topics for each emotional category. Subsequently, we employed logistic regression models to identify which topics significantly predict the presence of a specific emotion. For each respondent d, the model utilizes the continuous topic proportions $\theta_{d,k}$ as the sole predictors to output a predicted probability that a specific emotion E is present in the text. The relationship is formalized as follows:

$$\text{logit}(P(E_d = 1)) = \beta_0 + \sum_{k \in \mathcal{S}} \beta_k \theta_{d,k}$$

where $\mathcal{S}$ represents the subset of topics (e.g., $Topic_A, Topic_B, Topic_C$) identified as significant

predictors for the emotion E.

To evaluate the diagnostic accuracy of these predictions, we calculated the Area Under the ROC Curve (AUC).

The AUC measures the model's ability to distinguish between classes (e.g., "fear" vs. "no fear") by evaluating the trade-off between Sensitivity and Specificity across all possible classification thresholds. In this context:

- Sensitivity (True Positive Rate): The model's ability to correctly identify comments that actually contain the emotion.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

- Specificity (1 - False Positive Rate): The model's ability to correctly ignore comments that do not contain the emotion.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

The AUC is formally defined as the integral of the Receiver Operating Characteristic (ROC) curve, representing the probability that the model ranks a random positive instance higher than a random negative one:

$$\text{AUC} = \int_0^1 \text{Sensitivity}(\text{Specificity}^{-1}(t))dt$$

An AUC value of 0.5 indicates performance no better than random chance, while values closer to 1.0 represent superior discriminatory power. In this study, an increase in AUC when moving from a single topic to a cluster indicates that the combined topic proportions provide a more accurate and robust emotional profile than individual themes.

The final metric employed to evaluate the relationship between themes and emotions is Cohen's d. This is a measure of effect size used to quantify the standardized difference between the means of two groups. It is calculated by dividing the difference between the group means (e.g., the mean proportion of Topic 13 in "sad" vs. "non-sad" texts) by the pooled standard deviation of the dataset.

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_p}$$

Where:

- $\bar{x}_1$ and $\bar{x}_2$ are the mean topic proportions for the two groups.
- s_1^2 and s_2^2 are the respective variances.
- n_1 and n_2 are the sample sizes for each group.

Values around 0.2 indicate a small effect, 0.5 a medium effect, and 0.8 or higher a large effect. This metric provides insight into the practical significance of the relationship, beyond mere

statistical p-values.

Sadness is significantly predicted by a cluster composed of Topic 9 (Inability to meet living standards) and Topic 13 (Struggle to make ends meet). While Topic 13 alone yielded a Cohen's d of -0.2339 and an AUC of 0.5566, the combined model reached a Cohen's d of -0.4211 and an AUC of 0.6373. This comparison demonstrates that the two topics considered together have a significantly higher discriminatory power. These results suggest that sadness in the context of precariousness is more accurately identified when these two closely related topics of material strain are accounted for simultaneously.

For *Disappointment*, the most relevant predictors are Topic 14 (Lack of resources), Topic 13, and Topic 6 (Illegal and unsafe working conditions). The multivariate model produced a combined Cohen's d of -0.3741 and an AUC of 0.611, representing a modest improvement over the best individual predictor, Topic 14 (AUC: 0.574; Cohen's d: -0.200). *Disapproval* is primarily characterized by Topic 3 (Poor working conditions), which emerged as the sole significant predictor in the logistic regression. This association resulted in a Cohen's d of -0.3345 and an AUC of 0.601. The *Neutral* category is significantly associated with Topic 5 (Overtime without proper compensation) and Topic 18 (Low wages). The combined model yielded a Cohen's d of -0.32 and an AUC of 0.58, indicating limited discriminatory capacity.

Across these three emotional categories, the effect sizes and AUC values remain in a moderate range, indicating that while the associations are statistically reliable, these emotions are thematically diffuse and not tightly concentrated around a single structural condition.

In contrast, *Fear* shows the strongest effect sizes among the emotions considered. It is predicted by a combination of Topic 2 (Unaffordable independent living), Topic 1 (Resource-driven economic insecurity), and Topic 3. This cluster achieved a Cohen's d of -1.4242 and an AUC of 0.7928, substantially higher than the corresponding single-topic model. However, this result must be interpreted with caution, as fear represents a very small proportion of the corpus. The limited number of cases may inflate effect sizes and reduce model stability. While the findings suggest that fear is closely associated with existential threats related to housing autonomy and structural economic insecurity, further evidence on larger samples would be necessary to confirm the robustness of this pattern.

7. Conclusions

This thesis has investigated the subjective experience of precariousness among Spanish youth by applying advanced Natural Language Processing techniques to open-ended survey responses from the *VulnYouth* study. While traditional sociological research often relies on predefined indicators of job insecurity, this study aims to uncover the latent themes and emotional undercurrents that young people use to define their own condition by integrating Structural Topic Modeling (STM) with Transformer-based emotion detection (BERT and RoBERTa).

A primary finding of this research is the specific emotional signature of precariousness in Spain. Contrary to the expectation that high job insecurity might fuel *Anger* (0.5%) or *Fear* (1.6%), our categorical emotion analysis (using RoBERTa) reveals a landscape dominated by *Disappointment* (51.2%) and *Sadness* (48.8%).

The high prevalence of *Disappointment*—defined as the reaction to unfulfilled expectations—suggests that young Spaniards view precariousness not as a temporary struggle, but as a broken social contract. They do not describe their situation with the active arousal of conflict, but with the low-dominance vocabulary of resignation. The dimensional analysis (VAD) confirms this, showing a distribution of scores skewed toward low *Dominance*, indicating a sense of powerlessness and a lack of agency over one's life trajectory.

By aggregating the 28 emotional labels into positive, negative, mixed, and neutral categories, this study identifies a strange sociological phenomenon: the Neutrality Paradox. Logistic regression analysis reveals that respondents facing the highest levels of objective job insecurity and economic strain are significantly more likely to adopt neutral, detached language rather than employing emotional descriptor.

This counter-intuitive finding challenges the assumption that greater vulnerability leads to more intense emotional expression. Instead, it suggests a mechanism of "affective numbing" or normalization: those living in the most acute precarity may view their condition as a factual reality to be described rather than a grievance to be expressed.

Ultimately, the Structural Topic Model identified 20 latent themes that confirm the multidimensional nature of precariousness. To move beyond descriptive mapping, this thesis systematically examined whether specific emotions are statistically associated with the probability of discussing particular themes.

We find out that *Sadness* is significantly associated with a cluster composed of Topic 9 (Inability to meet living standards) and Topic 13 (Struggle to make ends meet). *Disappointment* emerges in connection with structural deprivation and unsafe conditions (Topics 14, 13, and 6), *Disapproval* is primarily anchored in Topic 3 (Poor working conditions) and *Neutral* is associated with Topic 5 (Overtime without proper compensation) and Topic 18 (Low wages). Across these three emotional categories, effect sizes and AUC values remain in a moderate range, suggesting that while the associations are statistically reliable, these emotions are thematically diffuse rather than tightly concentrated around a single dominant dimension of precarity.

In contrast, Fear displays the strongest effect sizes and highest discriminatory power when predicted by a combination of Topic 2 (Unaffordable independent living), Topic 1 (Resource-driven economic insecurity), and Topic 3. However, this result must be interpreted cautiously: fear represents a very small proportion of the corpus, and the limited number of cases may inflate effect sizes and reduce model stability. While the pattern points toward a concentrated emotional response to existential threat, further research with larger samples would be required to confirm its robustness.

Methodological Contributions

From a methodological perspective, this thesis demonstrates the superiority of modern Transformer architectures over traditional lexicon-based approaches for sociological text analysis. The fine-tuned RoBERTa model detected nuanced emotional categories in short, highly sparse survey responses where traditional methods often fail. Furthermore, the use of STM with covariates allowed us to test the stability of these themes across gender and class lines, revealing a remarkable homogeneity in how precariousness is conceptualized across the demographic spectrum.

A promising avenue for future research involves the application of Spanish-native Large Language Models directly to the original untranslated responses. While the current study successfully leveraged the advanced capabilities of English-native Transformers through translation, a direct analysis of the original Spanish text could provide a critical cross-lingual validation.

In conclusion, this thesis defines youth precariousness in Spain as a pervasive emotional state of disappointed resignation. It is a condition defined less by the fear of what might happen, and more by the sadness for a future that is perceived as already lost.

Acknowledgements

I would like to thank sociologist Vincenzo Favara for his contribution to Section 6. Our exchange of ideas was instrumental in framing the study in light of the results and in their qualitative interpretation.

References

[1] Joan Benach et al. Precarious employment and health inequalities: a research agenda. *The Lancet Public Health*, 8(10):e756–e757, 2023.

[2] Mireia Bolíbar, Francisco Linares, and José Martí. The impact of temporary employment on poverty and social exclusion in spain: A dual labor market perspective. *European Societies*, 21(5):733–757, 2019.

[3] Sven Buechel and Udo Hahn. Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 578–585, Valencia, Spain, 2017. Association for Computational Linguistics.

[4] Sven Buechel and Udo Hahn. Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 578–585, Valencia, Spain, 2017. Association for Computational Linguistics.

[5] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. Goemotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online, 2020. Association for Computational Linguistics.

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, 2019. Association for Computational Linguistics.

[7] Margalida Gili, Miquel Roca, Sanjay Basu, Martin McKee, and David Stuckler. The mental health risks of economic crisis in spain: evidence from primary care centres, 2006 and 2010. *European Journal of Public Health*, 23(1):103–108, 2013.

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, 2016.

[9] IGOP, Institut de Govern i Polítiques Públiques. Vulnyouth project, 2023.

[10] Annie Irvine and Nikolas Rose. Job insecurity and mental health. *Social Theory & Health*, 20:221–240, 2022.

[11] Daniel Jurafsky and James H Martin. *Speech and Language Processing*. Pearson, 3rd draft edition, 2024. Available at https://web.stanford.edu/ jurafsky/slp3/.

[12] Jochen Kleres. Emotions and narrative analysis: A methodological approach. *Journal for the Theory of Social Behaviour*, 41(2):182–202, 2010.

[13] Elena Di Liberatore. Precariousness and mental health among spanish youth. Master's thesis, Politecnico di Milano, 2024. Advisor: Alessandra Guglielmi, Co-advisor: Lara Maestripieri.

[14] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[15] L. Maestripieri, A. Lanau, R. Soler-i Martí, and M. Acebillo-Baqué. Intertwined precariousness and precarity: Disentangling a phenomenon that characterizes spanish youth. *International Journal of Social Welfare*, 2024.

[16] Abdullah Al Maruf, Fahima Khanam, Md. Mahmudul Haque, Zakaria Masud Jiyad, M. F. Mridha, and Zeyar Aung. Challenges and opportunities of text-based emotion detection: A survey. *IEEE Access*, 12:18416–18450, 2024.

[17] David Mimno, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. Optimizing semantic coherence in topic models. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 262–272, 2011.

[18] Saif M. Mohammad. A practical guide to sentiment annotation: Challenges and solutions. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 174–179, San Diego, California, 2016. Association for Computational Linguistics.

[19] Saif M. Mohammad. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184, 2018.

[20] Pansy Nandwani and Rupali Verma. A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1):81, 2021.

[21] Vikram Patel, Alan J. Flisher, Sarah Hetrick, and Patrick McGorry. Mental health of young people: a global public health challenge. *The Lancet*, 369(9569):1302–1313, 2007.

[22] Rukhma Qasim, Waqas Haider Bangyal, Mohammed A. Alqarni, and Abdulwahab Ali Almazroi. A fine-tuned bert-based transfer learning approach for text classification. *Journal of Healthcare Engineering*, 2022:1–17, 2022. Article ID 3498123.

[23] Margaret E. Roberts, Brandon M. Stewart, and Edoardo M. Airoldi. A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111(515):988–1003, 2016.

[24] Margaret E. Roberts, Brandon M. Stewart, and Dustin Tingley. stm: An r package for structural topic models. *Journal of Statistical Software*, 91(2):1–40, 2019.

[25] Margaret E. Roberts, Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4):1064–1082, 2014.

[26] Torkel Rönnblad, Erik Grönholm, Johanna Jonsson, Isa Koranyi, Cecilia Orellana, Bertina Kreshpaj, Lingjing Chen, Leo Stockfelt, and Theo Bodin. Precarious employment and mental health: a systematic review and meta-analysis of longitudinal studies. *Scandinavian Journal of Work, Environment & Health*, 45(5):429–443, 2019.

[27] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[28] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008. Curran Associates, Inc., 2017.

[29] Giulio Venturini. Bayesian modelling of mental health in young people explained by labour precariousness. Master's thesis, Politecnico di Milano, 2024. Advisor: Alessandra Guglielmi, Co-advisor: Lara Ivana Maestripieri.

Appendix A

This appendix provides detailed descriptions of the variables and structures of the three datasets utilized in this thesis: the primary survey dataset (VulnYouth) and the two external datasets used for fine-tuning the Transformer models (GoEmotions and EmoBank).

Table 5 details the socio-demographic and economic variables extracted from the VulnYouth dataset that were actively used in the structural topic modeling (STM) and logistic regression analyses. It describes their original format and the specific transformations applied to convert them into binary dummy covariates for the models.

Table 5: Dataset Variables and Applied Transformations

Original Name	Description	Original Categories	Transformation Applied
Gender	Respondent's gender	1: Man; 0: Not man	Used as a categorical predictor in logistic regression.
Ageyear	Age group	1: $<$ 30 years old; 0: $\geq$ 30	Converted to `age_dum` for logistic regression.
Migrant	Migration background	1: No migration; 2: Internal migrant; 3: Migrant background	Converted to `migr_dum` for logistic regression.
Jobinsec	Likelihood to lose job in next 12 months	1: Very likely; 2: Likely; 3: Unlikely; 4: Very unlikely	Recoded to `dum_jobinsec`: 1 if "Likely/Very Likely", 0 otherwise.
Normepres	Perceived social pressure/vulnerability	Continuous normalized score from 0 to 20	Recoded to `dum_normepres`: 1 if $>$ 0.8 (acute precarity), 0 otherwise.
Ecoind	Economic independence from family	5-point Likert scale	Recoded to `dum_ecoind`: 1 for highest category (fully independent), 0 otherwise.
Endmeet	Difficulty making ends meet	Scale from 1 (great difficulties) to 6 (very easily)	Recoded to `dum_endmeet`: 1 for "Great difficulty" or "Difficulty", 0 otherwise.
Unexpmix	Capacity to face a €700 unexpected expense	0: No; 1: Yes, autonomously; 2: Yes, with household	Recoded to `dum_unexpmix`: 1 if autonomously able, 0 otherwise.

Table 5: Summary of VulnYouth variables used in the computational models and their respective transformations.

Table 6 and Table 7 detail the structures and target variables of the two external English-language datasets used to perform supervised fine-tuning on the RoBERTa and BERT architectures.

Name	Meaning	Categories
Text	Text of the analyzed comment/response	String
Emotion	28 fine-grained emotion categories	1: Present; 0: Not present (Multi-label classification)
Labels	The specific emotional states detected	Admiration, Amusement, Anger, Annoyance, Approval, Caring, Confusion, Curiosity, Desire, Disappointment, Disapproval, Disgust, Embarrassment, Excitement, Fear, Gratitude, Grief, Joy, Love, Nervousness, Optimism, Pride, Realization, Relief, Remorse, Sadness, Surprise, Neutral

Table 6: GoEmotions Dataset Variables (used for Categorical Classification).

Name	Meaning	Categories
Text	Text of the analyzed sentence/response	String
Valence	Degree of pleasantness or unpleasantness (Positive vs. Negative)	Continuous scale (typically 1 to 5)
Arousal	Physiological and psychological intensity (from calm/resignation to excited/anger)	Continuous scale (typically 1 to 5)
Dominance	Perceived degree of control or agency the speaker feels over the situation	Continuous scale (typically 1 to 5)

Table 7: EmoBank Dataset Variables (used for Dimensional Regression).