Task 2

For our hackathon project, we utilized the Mental Health in Tech Survey dataset from Kaggle. Our hypothesis is that tendencies to "betray" a company may correlate with mental health challenges faced by employees. This concept can be framed within a military context: disgruntled tech workers can be likened to soldiers who might feel compelled to betray their organization due to unresolved mental health issues.

Rationale

1. **Relevance of Mental Health**: In a military setting, soldiers facing mental health challenges might experience increased stress, leading to feelings of disillusionment or betrayal toward their mission or leadership. This connection is crucial as it underscores the importance of mental health in maintaining loyalty and morale.
2. **Scalability of Approach**: Conducting a mental health survey can be more palatable for soldiers compared to a direct inquiry into betrayal tendencies. By focusing on mental health, we create an environment where individuals are more likely to provide honest feedback, which can be crucial for understanding their motivations and challenges.
3. **Practical Implications**: Understanding the mental health landscape among soldiers can inform leadership about the necessary actions to take, and which soldiers to target or keep an eye on.

Dataset Description

The dataset includes various numerical features such as:

- **Gender**
- **Family Mental Health History**
- **Therapy Seeking Behavior**
- **Relationships with Superiors/Comrades**

Additionally, it contains comments reflecting the interviewees' general thoughts on their mental health. To adapt the dataset for a military context, features that seemed out of place were mapped to more appropriate terms. For example, **work_interference** was changed to **mission_interference**.

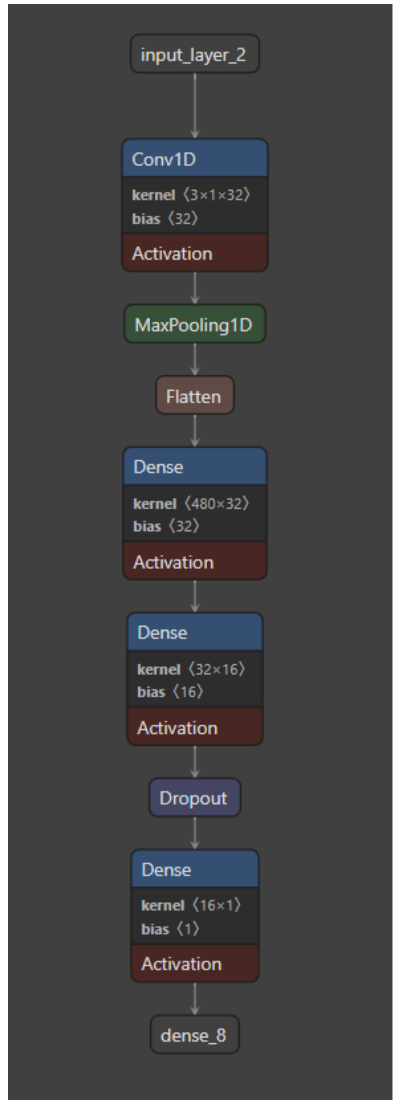
Synthetic Target Variable : Betrayal

- Since the dataset lacked a direct target variable to represent betrayal tendencies, we created a **synthetic variable** using key features from the data. The idea is that a poor score in certain areas, such as **relationships with comrades** or **mental and physical job consequences**, signals a higher likelihood of betrayal. Conversely, high scores in features like **frequency of leaves** and **mission interference** indicate job dissatisfaction, contributing to a betrayal risk.
- To build this variable, we classified each data point based on whether it had high scores in betrayal-related features (top 20%) and low scores in loyalty-related features (bottom 20%). These were then combined into a final betrayal score. If this score indicated a high likelihood of betrayal (over 50%), the employee was labeled as a potential risk (1). Otherwise, they were labeled as loyal (0).
- This variable allows us to capture the probability of betrayal, to use as a target variable for our model.

Modeling Approach

Our modeling approach involves generating probability arrays, where each element corresponds to the likelihood of a data point belonging to a specific class.

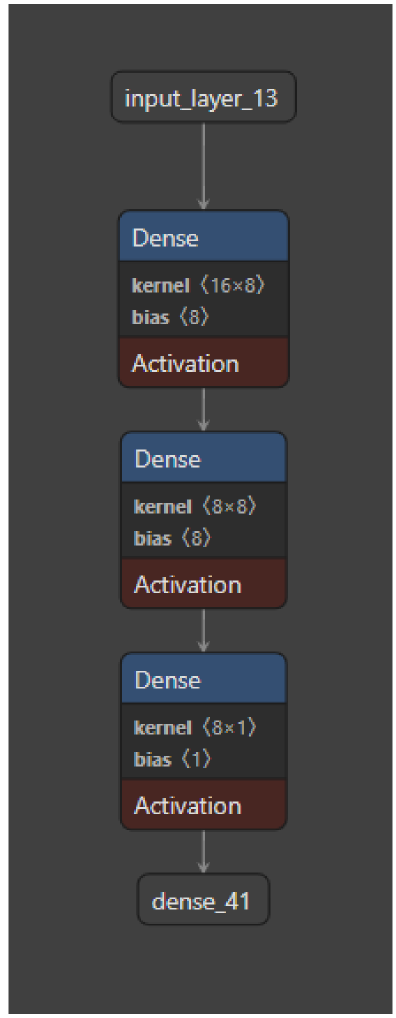
1. **Gaussian Naive Bayes Classifier**: We utilized this classifier for the numerical features of the dataset, which assumes that the features follow a Gaussian distribution. This model is particularly effective for high-dimensional data.
2. **Textual Data Processing**: The comments were vectorized using **TF-IDF vectorization**, which transforms the text into a numerical representation reflecting the importance of words in context. The vectorized data was then fed into a **1-Dimensional Convolutional Neural Network (CNN)**, using Binary Cross-Entropy Loss as our criterion. The CNN outputted probability arrays. The architecture of the CNN is described on the left.



Architecture of the TextCNN used.

Neural Network for Classification.

- The CNN mentioned above returns **1** if it predicts the soldier is disloyal and **0** if the soldier is loyal. This output is incorporated as an additional feature alongside various numerical features related to mental health and workplace dynamics. These features are then fed into the final neural network classifier, which synthesizes the information to classify the soldier as either **loyal** or a **risk**. By combining insights from both textual and numerical data, the model provides a robust assessment of loyalty.
- The architecture of the classifier consists of:
  - **Input Layer**: Accepting a total of 16 features, including the binary output from the CNN.
  - **Hidden Layers**: Two dense layers with 8 neurons each, using the ReLU activation function to introduce non-linearity and enhance learning capabilities.
  - **Output Layer**: A single neuron with a sigmoid activation function, which outputs a probability score indicating the likelihood of the soldier being classified as a risk (not loyal) versus loyal.



Considerations for Improvement :

1. **Data Granularity**:

   Much of the data in our analysis is binary (1/0). Implementing a rating system would provide a more nuanced and robust assessment of loyalty and betrayal. In practical applications, we would solicit soldiers for ratings on an arbitrary scale to capture a wider range of sentiments and experiences.

2. **Comment Reliability**:

   The comments included in the dataset are general observations and should not be blindly relied upon to accurately reflect loyalty or betrayal. For a more insightful analysis, we could ask soldiers targeted questions that delve deeper into their feelings and intentions. For instance, a more probing question could be: “How often do you feel that your well-being is compromised by your duties?” This could reveal underlying frustrations and potential motivations for disloyalty that mere passing comments might not capture.

Conclusion

In conclusion, we developed a robust model to assess soldier loyalty by integrating a Convolutional Neural Network (CNN) for textual analysis with a feedforward neural network classifier. The CNN processes comments to predict loyalty, generating binary outputs that, alongside key numerical features related to mental health and workplace dynamics, form inputs for the final classifier, which is a feed-forward neural network.