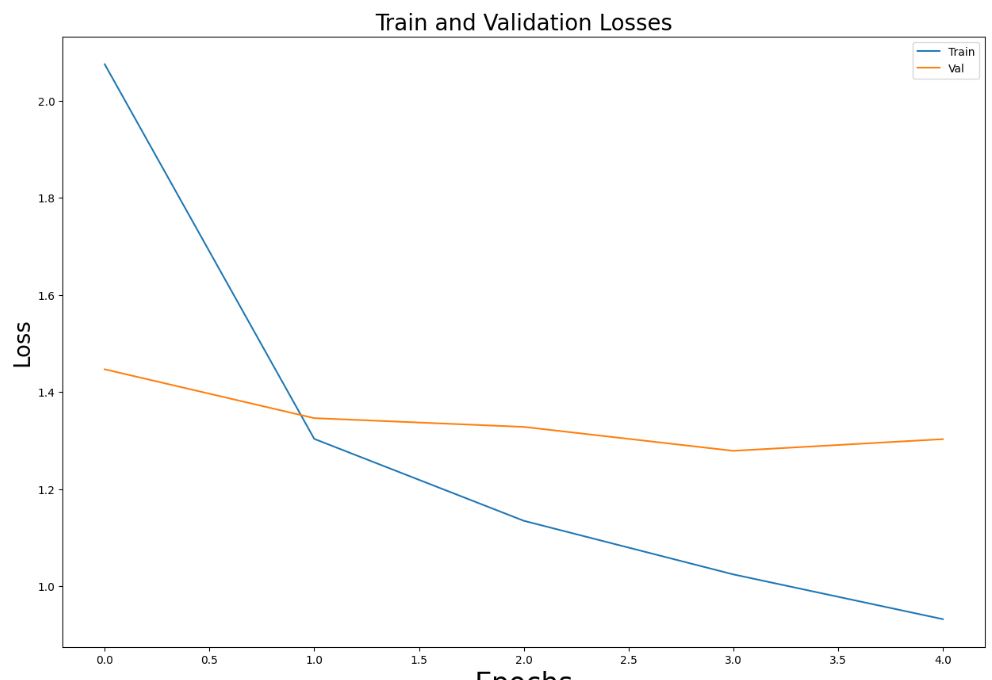
Distilbert

Learning rate: 3e-5

Epoch: 5

Batch Size = 16  
Full data

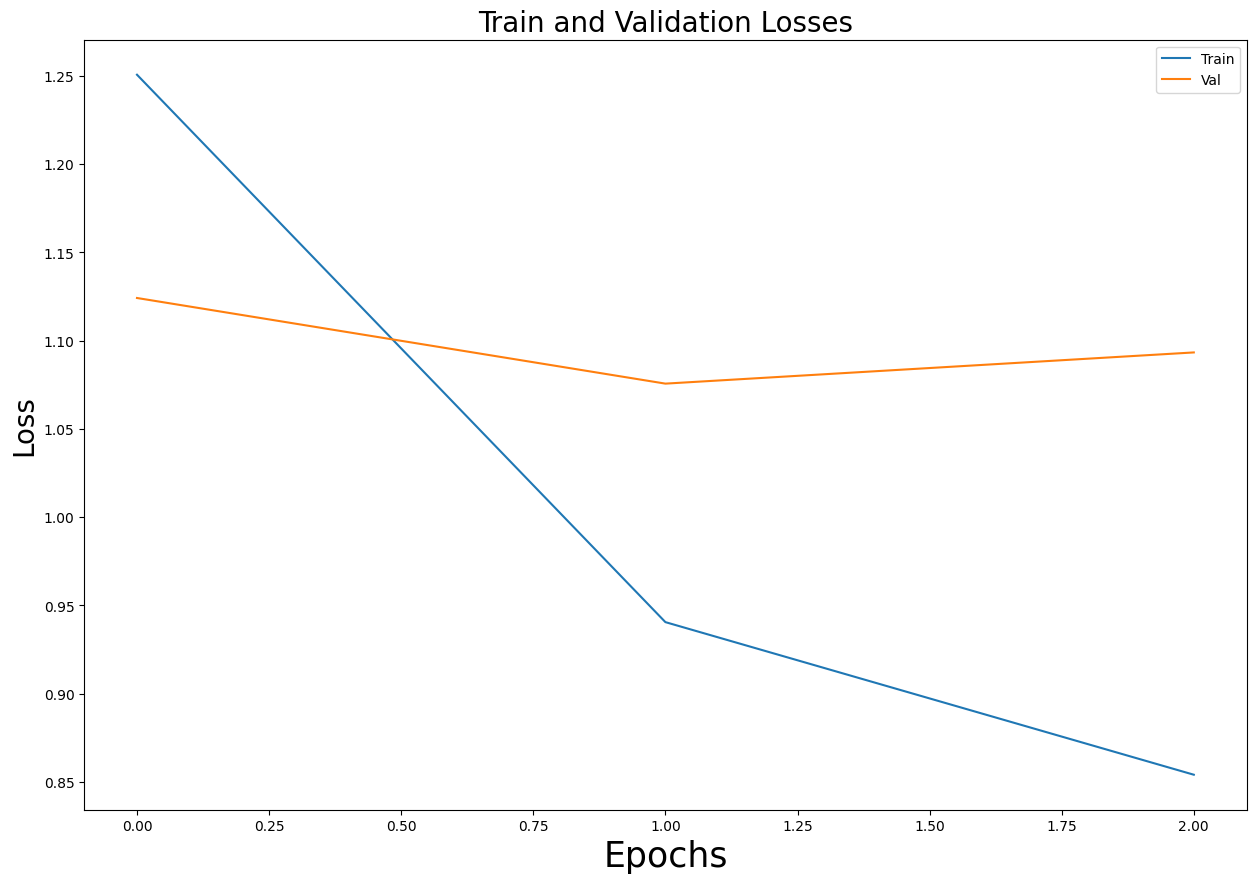


Albert

Learning rate: 3e-5

Epoch: 5

Batch Size = 16



Discussion and Future Work

In this study, we employed ALBERT, DistilBERT, and DistilRoBERTa models on both 50% and 100% of the training set data to assess their performance in a general domain context. While these models exhibited promising results, it's important to acknowledge that several factors influenced our choices and potential areas for improvement and expansion.

Model Selection and Limitations

We opted for ALBERT, DistilBERT, and DistilRoBERTa due to their efficiency and strong performance in various natural language processing tasks. However, it's worth noting that we did not include BERT in our experimentation due to limitations in time and computational resources. Future studies could incorporate BERT to provide a comprehensive comparison across a wider range of transformer-based models.

Furthermore, our models' fine-tuning could have been more efficient if we had explored various hyperparameters and tuning strategies. Investigating the impact of learning rates, batch sizes, and other parameters could potentially lead to improved results and better convergence rates.

Future Directions and Expansion

Several avenues for future research and exploration emerge from this study's findings:

* Specialized Domains: While our focus was on a general domain, future work could delve into specialized domains such as medical, legal, financial, or news media. Adapting pre-trained models to cater to domain-specific terminology and nuances could greatly enhance their effectiveness in those areas.
* Multilingual Analysis: The models' language capabilities could be evaluated and fine-tuned for specific languages. This is particularly relevant for languages with limited available data, as well as for cross-lingual applications.
* Fine-tuning Strategies: Investigating more advanced fine-tuning strategies, such as gradient accumulation, layer-wise adaptation, or transfer learning, could lead to even better performance and quicker convergence.
* Ensemble Approaches: Combining predictions from multiple models, especially from diverse architectures, could potentially enhance overall performance and robustness.
* Handling Domain Shift: Future studies should address the challenge of domain shift, where the model's performance might deteriorate when applied to data from different distributions or sources. Domain adaptation techniques could be explored to mitigate this issue.
* Model Interpretability: Enhancing the interpretability of transformer models remains a critical research area. Developing techniques to understand and visualize the reasoning behind model predictions can greatly enhance user trust and model adoption.
* Resource-Efficient Pre-training: Investigating methods to reduce the computational and memory requirements of pre-training can make these models more accessible and applicable to a wider range of researchers and practitioners.
* In conclusion, this study serves as a stepping stone in understanding the performance of ALBERT, DistilBERT, and DistilRoBERTa models in a general domain scenario. The choices made during this research open up a plethora of opportunities for future studies to explore model variants, parameter tuning, specialized domains, and language-specific adaptations. The advancements in natural language processing continue to evolve rapidly, and we believe that our findings pave the way for exciting developments and applications in the field.