

| Data Type | Model | Class | Layer 1 | | | | Layer 2 | | | | Layer 3 | | | | Layer 4 | | | | Layer 5 | | | |
|--------------------|---------------------------------------|------------|-----------|--------|---------|---------------|-----------|--------|---------|---------------|-----------|--------|---------|---------------|-----------|--------|---------|---------------|-----------|--------|---------|---------------|
| | | | Precision | Recall | F-score | ROC AUC score | Precision | Recall | F-score | ROC AUC score | Precision | Recall | F-score | ROC AUC score | Precision | Recall | F-score | ROC AUC score | Precision | Recall | F-score | ROC AUC score |
| Balanced Dataset | KoSimSE-suberta-multitask | Relevant | 0.786 | 0.691 | 0.735 | 0.729 | 0.813 | 0.672 | 0.736 | 0.743 | 0.802 | 0.725 | 0.762 | 0.764 | 0.776 | 0.722 | 0.748 | 0.749 | 0.791 | 0.664 | 0.722 | 0.728 |
| | | Irrelevant | 0.941 | 0.969 | 0.950 | 0.951 | 0.937 | 0.969 | 0.953 | 0.953 | 0.948 | 0.964 | 0.956 | 0.956 | 0.947 | 0.957 | 0.952 | 0.952 | 0.936 | 0.965 | 0.950 | 0.951 |
| | ko-suberta-multitask | Relevant | 0.815 | 0.696 | 0.751 | 0.756 | 0.758 | 0.681 | 0.717 | 0.720 | 0.774 | 0.731 | 0.752 | 0.753 | 0.884 | 0.591 | 0.708 | 0.738 | 0.752 | 0.591 | 0.662 | 0.672 |
| | | Irrelevant | 0.940 | 0.967 | 0.953 | 0.954 | 0.940 | 0.956 | 0.948 | 0.948 | 0.940 | 0.957 | 0.953 | 0.953 | 0.844 | 0.957 | 0.697 | 0.901 | 0.941 | 0.955 | 0.948 | 0.948 |
| | KR-SBERT-V40K42a/NLI-avgSTS | Relevant | 0.763 | 0.726 | 0.744 | 0.745 | 0.773 | 0.745 | 0.759 | 0.759 | 0.814 | 0.762 | 0.787 | 0.788 | 0.824 | 0.714 | 0.765 | 0.769 | 0.799 | 0.717 | 0.736 | 0.758 |
| | | Irrelevant | 0.956 | 0.954 | 0.955 | 0.955 | 0.950 | 0.950 | 0.950 | 0.950 | 0.954 | 0.963 | 0.958 | 0.959 | 0.957 | 0.964 | 0.950 | 0.951 | 0.944 | 0.965 | 0.952 | 0.954 |
| | distilase-bert-v2 | Relevant | 0.705 | 0.491 | 0.579 | 0.598 | 0.723 | 0.417 | 0.529 | 0.570 | 0.738 | 0.319 | 0.445 | 0.529 | 0.728 | 0.291 | 0.416 | 0.510 | 0.651 | 0.351 | 0.456 | 0.501 |
| | | Irrelevant | 0.908 | 0.957 | 0.932 | 0.933 | 0.897 | 0.960 | 0.927 | 0.929 | 0.970 | 0.974 | 0.919 | 0.922 | 0.877 | 0.974 | 0.925 | 0.926 | 0.904 | 0.966 | 0.929 | 0.930 |
| | paraphrase-unsiftlingual-MuSiM-L12-v2 | Relevant | 0.725 | 0.511 | 0.599 | 0.618 | 0.743 | 0.437 | 0.550 | 0.590 | 0.758 | 0.330 | 0.468 | 0.549 | 0.748 | 0.311 | 0.439 | 0.530 | 0.671 | 0.371 | 0.478 | 0.521 |
| | | Irrelevant | 0.901 | 0.950 | 0.925 | 0.926 | 0.899 | 0.962 | 0.929 | 0.931 | 0.869 | 0.973 | 0.918 | 0.921 | 0.879 | 0.976 | 0.925 | 0.928 | 0.898 | 0.970 | 0.933 | 0.934 |
| Imbalanced Dataset | KoSimSE-suberta-multitask | Relevant | 0.862 | 0.470 | 0.608 | 0.666 | 0.825 | 0.525 | 0.645 | 0.680 | 0.752 | 0.622 | 0.681 | 0.687 | 0.770 | 0.607 | 0.679 | 0.689 | 0.725 | 0.605 | 0.660 | 0.665 |
| | | Irrelevant | 0.960 | 0.994 | 0.977 | 0.977 | 0.968 | 0.983 | 0.975 | 0.976 | 0.970 | 0.981 | 0.975 | 0.976 | 0.940 | 0.995 | 0.967 | 0.968 | 0.974 | 0.980 | 0.977 | 0.977 |
| | ko-suberta-multitask | Relevant | 0.828 | 0.563 | 0.670 | 0.696 | 0.785 | 0.568 | 0.659 | 0.677 | 0.733 | 0.667 | 0.707 | 0.710 | 0.890 | 0.646 | 0.715 | 0.723 | 0.785 | 0.583 | 0.660 | 0.684 |
| | | Irrelevant | 0.965 | 0.990 | 0.977 | 0.978 | 0.965 | 0.987 | 0.976 | 0.976 | 0.970 | 0.980 | 0.975 | 0.975 | 0.972 | 0.987 | 0.979 | 0.980 | 0.965 | 0.986 | 0.975 | 0.976 |
| | KR-SBERT-V40K42a/NLI-avgSTS | Relevant | 0.791 | 0.474 | 0.591 | 0.633 | 0.841 | 0.536 | 0.684 | 0.709 | 0.738 | 0.649 | 0.699 | 0.704 | 0.859 | 0.598 | 0.705 | 0.729 | 0.785 | 0.523 | 0.628 | 0.654 |
| | | Irrelevant | 0.971 | 0.973 | 0.972 | 0.972 | 0.958 | 0.988 | 0.973 | 0.973 | 0.970 | 0.982 | 0.976 | 0.976 | 0.956 | 0.987 | 0.971 | 0.972 | 0.959 | 0.987 | 0.973 | 0.973 |
| | distilase-bert-v2 | Relevant | 0.764 | 0.312 | 0.443 | 0.538 | 0.702 | 0.333 | 0.452 | 0.518 | 0.737 | 0.294 | 0.420 | 0.516 | 0.778 | 0.266 | 0.448 | 0.547 | 0.766 | 0.239 | 0.387 | 0.513 |
| | | Irrelevant | 0.947 | 0.990 | 0.968 | 0.969 | 0.948 | 0.988 | 0.968 | 0.968 | 0.943 | 0.991 | 0.966 | 0.967 | 0.943 | 0.991 | 0.966 | 0.967 | 0.941 | 0.992 | 0.966 | 0.967 |
| | paraphrase-unsiftlingual-MuSiM-L12-v2 | Relevant | 0.843 | 0.333 | 0.475 | 0.587 | 0.764 | 0.357 | 0.487 | 0.561 | 0.774 | 0.326 | 0.459 | 0.550 | 0.856 | 0.335 | 0.485 | 0.596 | 0.823 | 0.272 | 0.409 | 0.548 |
| | | Irrelevant | 0.957 | 0.958 | 0.957 | 0.958 | 0.958 | 0.963 | 0.960 | 0.961 | 0.953 | 0.981 | 0.967 | 0.967 | 0.953 | 0.955 | 0.954 | 0.954 | 0.951 | 0.952 | 0.951 | 0.952 |

Table 1: Model Performance Across Different Layers on Balanced and Imbalanced dataset