**Table 1: Existing works on dialect classification**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| Authors | Regional Dialect | Data Source | Classifier/Model | **Dataset** | Reported Performance |
| (Faria et al., 2023)[13] | Bengali(Chittagong, Noakhali, Barisal, Sylhet and Mymensingh) | Websites, social media platforms, and discussion boards. | mBERT and Bangla-bert-base | Balanced | Accuracy-mBERT- 84.36%, Bangla-bert-base-85.86% |
| (El-Haj et al., 2018)[18] | Arabic (Egypt, North Africa, Gulf, Levant and Standard Arabic) | Arabic Commentary  Dataset (AOC) (Zaidan and Callison-Burch, 2014).[10] | Naïve Bayes, Support Vector, Machine (SVM), k–Nearest Neighbor (KNN) and Decision  Trees (J48). | **Imbalanced** | accuracy-76% |
| Doostmohammadi and Nassajian[19] | Akkadian(Old Babylonian, Mid-  dle Babylonian Peripheral, Standard Babylo-  nian, Neo-Babylonian, Late Babylonian, and  Neo-Assyrian.) | Jauhi-  ainen et al., 2019[20] | SVM, NF, RF, LR and Ensemble of SVM + naive Bayes | **Imbalanced** | Macro-  averaged F1-score 72.10%. |
| (Vaidya and Kane., 2023) [21] | English(American, British & Common), Spanish(European/Peninsular, Argentine & Common) and Portuguese (European, Brazilian & Common) | Dataset (Zampieri et al., 2023)[22] | BERT and RoBERTa | **Imbalanced** | Accuracy- 87.29% |
| (Abdelali et al., 2021) [23] | Arabic(Middle East  and North Africa). | QADI | AraBERT and mBERT. | **Imbalanced** | Accuracy-91.5% |
| Tamal et al., 2022[15] | Bengali ( Pabna, Chattagram) | local news portals,  dramas and social media | SVM, MNB, LR, RF, DT, and KNN. | **Somewhat Imbalanced** | Accuracy 96% |
|  |  |  |  |  |  |

**Table 2: Available dialectical datasets**

|  |  |  |  |
| --- | --- | --- | --- |
| Author | Dataset | Balanced | Class distribution Ratio |
| Zaidan and Callison-Burch, 2014[10] | Arabic On-line Commentary Data Set (AOC) | ❌ | 3.60 : 20.77 : 15.42 |
| Siewert et al[40] | Low Saxon Dialect Classification(LSDC) | ❌ | 1.11: 1.13: 1.33 ..... 71.07: 99.16: 487.31 |
| Alshutayri and Atwell[43] | Arabic dialect Text corpus | ❌ | 1.65 : 1.64 : 1.73 : 2.15 |
| Jihad Zahir[44] | Integrated Arabic dialect identification dataset(IADD) | ❌ | 2.58 : 13.11 : 18.10 : 35.03 : 405.43 |
| Barnes et al[45] | Norwegian dialect(NorDial) | ❌ | 1.95 : 1.26 : 7.06 |
| Camposampiero et al[41] | Italian dialects and language | ❌ | 1.42 : 1.44 : 1.57 ..... 6.23 : 6.81 : 12.11 |

**Table 3 Confusion Matrix**

|  |  |  |
| --- | --- | --- |
|  | Predictive Positive | Predictive Negative |
| Actual Positive | TP | FN |
| Actual Negative | FP | TN |

**TABLE 4 Classification models and their parameter configurations**

|  |  |  |
| --- | --- | --- |
| **Model** | **Overview** | **Parameters** |
|
| MNB | Multinomial Naive Bayes | Default |
| RF | Random Forest | n\_estimators=100, random\_state=1 |
| XGB | Extreme Gradient Boosting | objective='binary:logistic', random\_state=1 |
| DT | Decision Tree | random\_state=1 |
| LGBM | Light Gradient-Boosting Machine | random\_state=1 |
| LR | Logistic Regression | random\_state=1 |
| BNB | Bernoulli Naive Bayes | Default |

**TABLE 5 Python packages.**

|  |  |
| --- | --- |
| Python Package | Purpose |
| Scikit-learn | Implements machine learning models (e.g., DT, NB, RF, LR) and evaluation metrics (e.g., Precision, Recall, F1-Score, AUC, G-Mean). Provides utilities for TF-IDF feature extraction and data preprocessing. |
| imbalance-learn | Applies resampling techniques, including oversampling (e.g., SMOTE), undersampling (e.g., Random Undersampling), and hybrid methods (e.g., SMOTETomek) to handle imbalanced datasets. |
| Matplotlib & seaborn | Visualizes results, showcasing model performance and comparing metrics with and without resampling techniques. |
| Pandas & numpy | Facilitates dataset manipulation, transformation, and statistical computations, essential for handling tabular data. |

**TABLE 6: Data distribution before and after resampling.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Before sampling training data count | After sampling training data count | | | | | |
| Datasets | Dialects | Default | Smote | ROS | ADASYN | Tomek Links | SMOTE-  TOMEK | ROS-RUS |
| Bangla | Barisal | 719 | 3022 | 3022 | 2945 | 719 | 2981 | 2500 |
| Chittagong | 2741 | 3022 | 3022 | 2741 | 2741 | 3022 | 2500 |
| Mymensingh | 1988 | 3022 | 3022 | 1988 | 1988 | 3022 | 2500 |
| Standard\_Bangla | 748 | 3022 | 3022 | 748 | 748 | 3022 | 2500 |
| Sylhet | 3022 | 3022 | 3022 | 3022 | 2654 | 3022 | 2500 |
| Arabic | EGY | 3687 | 67906 | 67906 | 3687 | 3687 | 67714 | 5615 |
| GLF | 5362 | 67906 | 67906 | 5362 | 5362 | 67906 | 5615 |
| IRQ | 173 | 67906 | 67906 | 67923 | 173 | 67906 | 5615 |
| LEV | 67906 | 67906 | 67906 | 67906 | 66917 | 67906 | 5615 |
| MGH | 22726 | 67906 | 67906 | 22726 | 22726 | 67906 | 5615 |
| general | 1951 | 67906 | 67906 | 1951 | 1951 | 67906 | 5615 |
| Hindi | AWA | 9819 | 15774 | 15774 | 15859 | 9819 | 15597 | 19300 |
| BHO | 14983 | 15774 | 15774 | 14983 | 14983 | 15774 | 19300 |
| BRA | 15633 | 15774 | 15774 | 15633 | 15633 | 15774 | 19300 |
| HIN | 15446 | 15774 | 15774 | 15446 | 15446 | 15774 | 19300 |
| MAG | 15744 | 15774 | 15774 | 15744 | 15318 | 15774 | 19300 |

**Table 7: Average Impact Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | recall(pd) | AUC | G-mean | balance(bal) |
| Increased performance value | Positive impact (Maximum) | 2.00% Tomek Links-RF | 2.07% Tomek Links-RF | 3.86% SMOTETomek-XGB | 1.45% Tomek Links-RF |
| Positive  impact  (Minimum) | 0.10% Tomek Links-LGBM | 0.02% Tomeklinks-LR & SMOTETomek-LGBM | 0.16% Tomek Links-LGBM | 0.08% Tomeklinks-LGBM |
| Decreased performance value | Negative  impact  (Maximum) | 7.38%  ROS-RUS-LGBM | 2.47%  ROS-RUS-DT | 2.4  ROS-RUS-DT | 5.16%  ROS-RUS-DT |
| Negative  impact (Minimum) | 0.28% ADASYN-LGBM | 0.01% Tomeklinks-MNB & SMOTE-LGBM | 0.17 TomekLinks-DT | 0.15 ADASYN-LGBM |

**Table 8: Performance report of baseline models.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Recall (*pd*) | | | pf | | | AUC | | | G-mean | | |
| AVG | MAX | MIN | AVG | MAX | MIN | AVG | MAX | MIN | AVG | MAX | MIN |
| MNB | 0.8414 | 0.8628 | 0.8082 | 0.0521 | 0.0662 | 0.0347 | 0.9666 | 0.9792 | 0.9574 | 0.8571 | 0.912 | 0.7996 |
| RF | 0.7999 | 0.839 | 0.7679 | 0.0601 | 0.0655 | 0.0521 | 0.9412 | 0.9543 | 0.9339 | 0.8362 | 0.8662 | 0.8027 |
| XGB | 0.7923 | 0.8074 | 0.7761 | 0.0688 | 0.0925 | 0.0521 | 0.9446 | 0.9602 | 0.9261 | 0.8087 | 0.8663 | 0.7174 |
| DT | 0.7345 | 0.8043 | 0.6928 | 0.0737 | 0.0798 | 0.0678 | 0.8144 | 0.8197 | 0.8112 | 0.7995 | 0.8088 | 0.7932 |
| LGBM | 0.8006 | 0.8377 | 0.7397 | 0.063 | 0.0751 | 0.0442 | 0.9524 | 0.9699 | 0.9404 | 0.8267 | 0.8871 | 0.772 |
| LR | **0.8573** | **0.8774** | **0.8321** | **0.0431** | **0.0479** | **0.0346** | **0.9701** | 0.9789 | **0.9654** | **0.8832** | **0.9121** | 0.8559 |
| BNB | 0.8322 | 0.8565 | 0.8165 | 0.0438 | 0.053 | 0.0357 | 0.9659 | **0.98** | 0.9523 | 0.882 | 0.9092 | **0.8661** |

**Table 9: Performance report of Oversampling techniques( SMOTE, ROS, and ADASYN)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Recall | | | pf | | | AUC | | | G-mean | | |
| AVG | MAX | MIN | AVG | MAX | MIN | AVG | MAX | MIN | AVG | MAX | MIN |
| MNB-SMOTE | 0.8214 | 0.861 | 0.7595 | 0.0398 | 0.0464 | 0.0346 | 0.9676 | 0.979 | 0.9461 | 0.8897 | 0.912 | 0.8549 |
| RF-SMOTE | 0.7781 | 0.7875 | 0.7644 | 0.0579 | 0.0626 | 0.0528 | 0.9747 | 0.9934 | 0.9602 | 0.8427 | 0.8643 | 0.8239 |
| XGB-SMOTE | 0.7596 | 0.7878 | 0.727 | 0.0576 | 0.0613 | 0.0528 | 0.9609 | 0.9681 | 0.9531 | 0.8418 | 0.8642 | 0.8185 |
| DT-SMOTE | 0.707 | 0.7413 | 0.6772 | 0.0764 | 0.0846 | 0.0705 | 0.8821 | 0.9513 | 0.83 | 0.7915 | 0.8067 | 0.7806 |
| LGBM-SMOTE | 0.7654 | 0.8193 | 0.7239 | 0.0542 | 0.0623 | 0.045 | 0.9677 | 0.9712 | 0.9612 | 0.8504 | 0.8848 | 0.827 |
| LR-SMOTE | 0.8203 | 0.8583 | 0.7702 | 0.041 | 0.0464 | 0.0351 | 0.9711 | 0.9794 | 0.9556 | 0.8869 | 0.9106 | 0.8559 |
| BNB-SMOTE | 0.8129 | 0.8531 | 0.7691 | 0.0413 | 0.0439 | 0.0362 | 0.9727 | 0.9854 | 0.9519 | 0.8859 | 0.9077 | 0.8625 |
| MNB-ROS | 0.8184 | 0.8612 | 0.7551 | 0.0402 | 0.047 | 0.0345 | 0.9735 | 0.9815 | 0.9592 | 0.8882 | 0.9121 | 0.8527 |
| RF-ROS | 0.7938 | 0.832 | 0.7584 | 0.0579 | 0.0628 | 0.0521 | **0.9796** | **0.9963** | 0.9646 | 0.8428 | 0.8662 | 0.8235 |
| XGB-ROS | 0.7457 | 0.7875 | 0.6996 | 0.058 | 0.0613 | 0.0528 | 0.963 | 0.9707 | 0.9565 | 0.839 | 0.8643 | 0.8114 |
| DT-ROS | 0.6996 | 0.7452 | 0.6503 | 0.0782 | 0.0902 | 0.0703 | 0.9132 | 0.9779 | 0.8475 | 0.7865 | 0.8071 | 0.7645 |
| LGBM-ROS | 0.7684 | 0.8185 | 0.7322 | 0.0529 | 0.0609 | 0.0451 | 0.9693 | 0.9715 | 0.9655 | 0.8539 | 0.8845 | 0.8345 |
| LR-ROS | 0.8292 | **0.862** | 0.7854 | **0.0393** | **0.0431** | **0.0344** | 0.9763 | 0.9822 | 0.9673 | **0.8922** | **0.9125** | 0.8666 |
| BNB-ROS | 0.8116 | 0.8527 | 0.7644 | 0.0414 | 0.0447 | 0.0362 | 0.9697 | 0.9823 | 0.9461 | 0.8854 | 0.9076 | 0.8601 |
| MNB-ADASYN | 0.8303 | 0.8599 | 0.8078 | 0.0469 | 0.0564 | 0.0348 | 0.9752 | 0.9834 | 0.9649 | 0.8719 | 0.9114 | 0.8307 |
| RF-ADASYN | 0.7916 | 0.8308 | 0.7575 | 0.0613 | 0.068 | 0.053 | 0.9661 | 0.9816 | 0.9553 | 0.8325 | 0.8636 | 0.7953 |
| XGB-ADASYN | 0.779 | 0.8052 | 0.7453 | 0.0694 | 0.0916 | 0.0531 | 0.967 | 0.9785 | 0.9599 | 0.8066 | 0.8635 | 0.7206 |
| DT-ADASYN | 0.7221 | 0.797 | 0.6711 | 0.0766 | 0.0852 | 0.0693 | 0.8626 | 0.9208 | 0.8259 | 0.7915 | 0.804 | 0.7783 |
| LGBM-ADASYN | 0.7963 | 0.8298 | 0.7406 | 0.0621 | 0.0754 | 0.0451 | 0.9722 | 0.9848 | 0.9598 | 0.8287 | 0.8846 | 0.7718 |
| LR-ADASYN | **0.8351** | 0.8581 | **0.8108** | 0.0448 | 0.0516 | 0.0351 | 0.9786 | 0.9862 | 0.971 | 0.8778 | 0.9106 | 0.8452 |
| BNB-ADASYN | 0.823 | 0.853 | 0.7926 | 0.043 | 0.0511 | 0.0362 | 0.9794 | 0.9846 | **0.9725** | 0.8834 | 0.9077 | **0.8686** |

**Table 10 : Performance report of Undersampling techniques (TomekLinks)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Recall | | | pf | | | AUC | | | G-mean | | |
| AVG | MAX | MIN | AVG | MAX | MIN | AVG | MAX | MIN | AVG | MAX | MIN |
| MNB-Tomeklinks | 0.8413 | 0.861 | 0.8091 | 0.0518 | 0.0661 | 0.0351 | 0.9693 | 0.9807 | 0.9585 | 0.8578 | 0.9108 | 0.8 |
| RF-Tomeklinks | 0.8066 | 0.8438 | 0.7831 | 0.0579 | 0.0632 | 0.052 | 0.9439 | 0.9559 | 0.9366 | 0.8424 | 0.8667 | 0.8099 |
| XGB-Tomeklinks | 0.7894 | 0.8079 | 0.7705 | 0.0694 | 0.0926 | 0.053 | 0.945 | 0.9616 | 0.9259 | 0.8072 | 0.864 | 0.717 |
| DT-Tomeklinks | 0.7312 | 0.8092 | 0.6811 | 0.0749 | 0.0844 | 0.0661 | 0.8106 | 0.8186 | 0.8019 | 0.7966 | 0.8067 | 0.7815 |
| LGBM-Tomeklinks | 0.8001 | 0.8391 | 0.7401 | 0.0631 | 0.0747 | 0.045 | 0.9536 | 0.9716 | 0.9421 | 0.8267 | 0.8851 | 0.7732 |
| LR-Tomeklinks | **0.8607** | **0.8798** | **0.8395** | **0.0416** | **0.0469** | **0.0345** | **0.9731** | 0.9803 | **0.9666** | **0.8871** | **0.9124** | 0.8591 |
| BNB-Tomeklinks | 0.8333 | 0.8569 | 0.8134 | 0.0435 | 0.0532 | 0.0356 | 0.9676 | **0.9816** | 0.9525 | 0.8828 | 0.9094 | **0.8653** |

**Table 11 : Performance report of Hybrid-Sampling (SMOTETomek, and ROS-RUS)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Recall | | | pf | | | AUC | | | G-mean | | |
| AVG | MAX | MIN | AVG | MAX | MIN | AVG | MAX | MIN | AVG | MAX | MIN |
| MNB-SMOTETomek | **0.8208** | **0.8606** | 0.7596 | **0.04** | 0.0464 | **0.0347** | 0.9675 | 0.98 | 0.9458 | **0.889** | **0.9116** | 0.8547 |
| RF-SMOTETomek | 0.7769 | 0.7875 | 0.7584 | 0.0576 | 0.0617 | 0.0528 | **0.9748** | **0.9934** | 0.9595 | 0.8431 | 0.8642 | 0.824 |
| XGB-SMOTETomek | 0.7613 | 0.7905 | 0.7281 | 0.0575 | 0.0614 | 0.0523 | 0.9618 | 0.9718 | 0.9529 | 0.8421 | 0.8657 | 0.818 |
| DT-SMOTETomek | 0.7093 | 0.7416 | 0.6859 | 0.0755 | 0.081 | 0.0707 | 0.8848 | 0.9517 | 0.8297 | 0.7938 | 0.8052 | 0.7867 |
| LGBM-SMOTETomek | 0.7663 | 0.8197 | 0.7223 | 0.0539 | 0.061 | 0.045 | 0.9688 | 0.9752 | **0.9608** | 0.8511 | 0.885 | 0.8256 |
| LR-SMOTETomek | 0.8195 | 0.8589 | **0.7757** | 0.0414 | 0.0455 | 0.035 | 0.9718 | 0.9812 | 0.9567 | 0.886 | 0.9109 | 0.859 |
| BNB-SMOTETomek | 0.8119 | 0.854 | 0.7696 | 0.0416 | **0.0449** | 0.036 | 0.9723 | 0.9853 | 0.9515 | 0.8852 | 0.9082 | **0.8628** |
| MNB-ROS-RUS | 0.802 | 0.8595 | 0.7161 | 0.0432 | 0.0533 | 0.035 | 0.9626 | 0.9791 | 0.9341 | 0.8787 | 0.9111 | 0.8309 |
| RF-ROS-RUS | 0.7475 | 0.7915 | 0.6969 | 0.0596 | 0.0637 | 0.052 | 0.9554 | 0.9633 | 0.9417 | 0.8352 | 0.8665 | 0.8012 |
| XGB-ROS-RUS | 0.7348 | 0.7884 | 0.6698 | 0.06 | 0.0653 | 0.0526 | 0.948 | 0.9602 | 0.924 | 0.8322 | 0.8648 | 0.7928 |
| DT-ROS-RUS | 0.6603 | 0.6988 | 0.6097 | 0.0807 | 0.0837 | 0.0751 | 0.8401 | 0.8594 | 0.8215 | 0.7743 | 0.8042 | 0.7369 |
| LGBM-ROS-RUS | 0.7507 | 0.8191 | 0.7042 | 0.0572 | 0.0679 | 0.045 | 0.9528 | 0.9698 | 0.9312 | 0.8415 | 0.8849 | 0.8155 |
| LR-ROS-RUS | 0.8189 | 0.8596 | 0.7571 | 0.0412 | 0.0484 | 0.035 | 0.9641 | 0.9788 | 0.9371 | 0.8859 | 0.911 | 0.8493 |
| BNB-ROS-RUS | 0.7903 | 0.8526 | 0.7093 | 0.0451 | 0.054 | 0.0363 | 0.961 | 0.9807 | 0.9266 | 0.8732 | 0.9076 | 0.8281 |

**Table 12: Summary of best performing prediction results ( Recall range from 83% to 86.05%)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Recall | | | pf | | | AUC | | | G-mean | | |
| AVG | MAX | MIN | AVG | MAX | MIN | AVG | MAX | MIN | AVG | MAX | MIN |
| MNB | 0.8396 | 0.8605 | 0.8048 | 0.0528 | 0.0666 | 0.0352 | 0.9666 | 0.9792 | 0.9574 | 0.8553 | 0.9105 | 0.7986 |
| LR | 0.8537 | **0.8806** | 0.8178 | 0.044 | 0.0503 | **0.0345** | 0.9701 | 0.9789 | 0.9654 | 0.8811 | **0.9124** | 0.8584 |
| MNB-ADASYN | 0.8303 | 0.8599 | 0.8078 | 0.0469 | 0.0564 | 0.0348 | 0.9752 | 0.9834 | 0.9649 | 0.8719 | 0.9114 | 0.8307 |
| LR-ADASYN | 0.8351 | 0.8581 | 0.8108 | 0.0448 | 0.0516 | 0.0351 | **0.9786** | **0.9862** | **0.971** | 0.8778 | 0.9106 | 0.8452 |
| MNB-Tomeklinks | 0.8413 | 0.861 | 0.8091 | 0.0518 | 0.0661 | 0.0351 | 0.9693 | 0.9807 | 0.9585 | 0.8578 | 0.9108 | 0.8 |
| LR-Tomeklinks | **0.8607** | 0.8798 | **0.8395** | **0.0416** | **0.0469** | **0.0345** | 0.9731 | 0.9803 | 0.9666 | **0.8871** | **0.9124** | 0.8591 |
| BNB-Tomeklinks | 0.8333 | 0.8569 | 0.8134 | 0.0435 | 0.0532 | 0.0356 | 0.9676 | 0.9816 | 0.9525 | 0.8828 | 0.9094 | **0.8653** |

**Table 13 Performance impact summary of Arabic, Bangla, and Hindi**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | recall(pd) | G-mean | AUC | balance(bal) |
| Bangla | Increased performance value | Positive impact (Maximum) | 3.90% (SMOTE-MNB) | 4.53%  (SMOTE-MNB) | 1.84% (TomekLinks-RF) | 3%  (SMOTE-MNB) |
| Positive  impact  (Minimum) | 0.04% (SMOTETomek-DT & ROS-RUS-RF) | 0.34%  (TomekLinks-LGBM) | 0.02%  (SMOTETomek-XGB) | 0.07% (SMOTETomek-DT) |
| Decreased performance value | Negative  impact  (Maximum) | -3.52%  (ROS-DT) | -1.92%  (ROS-DT) | -2.17% (ROS-DT) | -2.53% (ROS-DT) |
| Negative  impact (Minimum) | -0.31% (ADASYN-BNB) | -0.19%  (ROS-RUS-DT) | -0.02% (TomekLinks-MNB) | -0.07% (ADASYN-BNB) |
| Arabic | Increased performance value | Positive impact (Maximum) | 0.45%  (TomekLinks-RF) | 10.43%  (SMOTE-XGB) | 2.4%  (TomekLinks-RF) | 3% (TomekLinks-RF) |
| Positive  impact  (Minimum) | 0.01% (TomekLinks-MNB) | 0.05% (TomekLinks-RF) | 0.01% (TomekLinks-BNB) | 0.01% (TomekLinks-LGBM) |
| Decreased performance value | Negative  impact  (Maximum) | -20.24% (ROS-RUS-DT) | -6.66% (ROS-RUS-DT) | -6.79% (ROS-RUS-DT) | -10.15 (ROS-RUS-DT) |
| Negative  impact (Minimum) | -0.08% (TomekLinks-LR) | -0.05% (ADASYN-BNB & TomekLinks-BNB) | -0.01% (TomekLinks-LGBM & TomekLinks-LR) | -0.01% (ADASYN-XGB) |
| Hindi | Increased performance value | Positive impact (Maximum) | 2.60% (TomekLinks-RF) | 1.69% (TomekLinks-RF) | 1.96% (TomekLinks-RF) | 1.89% (TomekLinks-RF) |
| Positive  impact  (Minimum) | 0.01%  (SMOTETomek-MNB) | 0.01% (ROS-LR, ROS-RUS-XGB & ROS-RUS-LGBM) | 1.83% (ADASYN-RF & SMOTETomek-RF) | 0.01% (SMOTETomek-MNB) |
| Decreased performance value | Negative  impact  (Maximum) | -0.64%  (ADASYN-DT) | -0.36% (ADASYN-DT) | -0.35% (ADASYN-DT) | -0.45% (ADASYN-DT) |
| Negative  impact (Minimum) | -0.04% (SMOTETomek-XGB) | -0.05% (SMOTE-XGB & ROS-DT) | -0.01%  (ROS-LR,ADASYN-BNB,TomekLinks-XGB & SMOTETomek-MNB) | -0.02% (SMOTETomek-XGB) |