**Results and Discussion:**

**GloVe Results:**

**Apriori Results:**

**Machine Learning Results:**

**First Arrangement**

* The most common ADR (Pain) as labels.
* The most common disease (Hypertensive Disease) as labels.
* Coughing as the most common ADR in the AskAPatient dataset, however it will not be used in the other subsets.
* Counts for ADRs, diseases and mental issues as labels.
* Testing the best number of estimators in terms of F-Score for random forests was made by trying all estimators between 100 and 1000 at 100 increment.
* Mental issues mentions were ignored as labels because of they were very little mentions.

**Random Forests with SMOTENC:**

From the given charts, the following can be observed:

* Hyper Tension was successfully predicted with blood pressure, with weights/heights being second. This gives importance to both readings as they can predict certain diseases.
* For pain, weights/heights have the biggest effect, which could indicate that pain as an ADR is cause by weights gain in relation to height.
* It is much easier to detect the count of mental issues and disease compared to ADRs.
* AskAPatient dataset was more accurate in its predictions than the complete dataset, this can be attributed to the fact that the user ADR is 100% correct, with no offset for mistakes and no imputations to fill the gaps.
* Continuous user information in general carry a lot more weight to influence health predictions, and therefore they are the most important factor to be used.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Complete Subset | Best Estimator | Accuracy | Precision | Recall | FScore |
| Hypertensive disease | 600 | 0.592814371 | 0.593138665 | 0.592868526 | 0.592540027 |
| Pain | 100 | 0.666191155 | 0.665535147 | 0.664103673 | 0.664272031 |
| ADR Count | 700 | 0.448881789 | 0.455145388 | 0.448323148 | 0.449865295 |
| Disease Count | 400 | 0.494252874 | 0.489404455 | 0.488715998 | 0.482368668 |
| Mental Count | 800 | 0.616891065 | 0.61115795 | 0.617872777 | 0.612736515 |
| Weights Subset | Best Estimator | Accuracy | Precision | Recall | Fscore |
| Hypertensive disease | 400 | 0.729166667 | 0.739564428 | 0.729166667 | 0.7261957 |
| Pain | 300 | 0.796296296 | 0.807586207 | 0.827941176 | 0.794534763 |
| ADR Count | 100 | 0.454545455 | 0.508639911 | 0.452380952 | 0.45526457 |
| Disease Count | 800 | 0.655172414 | 0.657017544 | 0.64983165 | 0.651343256 |
| Mental Count | 600 | 0.639344262 | 0.64957265 | 0.655716586 | 0.635406699 |
| Pressure Subset | Best Estimator | Accuracy | Precision | Recall | Fscore |
| Hypertensive disease | 700 | 0.763636364 | 0.762952605 | 0.763262599 | 0.763079415 |
| Pain | 100 | 0.771144279 | 0.771237624 | 0.77156027 | 0.771093286 |
| ADR Count | 300 | 0.579710145 | 0.578879959 | 0.578689196 | 0.577274984 |
| Disease Count | 500 | 0.709401709 | 0.709569157 | 0.710602981 | 0.704658615 |
| Mental Count | 200 | 0.74025974 | 0.743861768 | 0.745692686 | 0.744245344 |
| Ask A Patient | Best Estimator | Accuracy | Precision | Recall | Fscore |
| Hypertensive disease | 200 | 0.720547945 | 0.718644629 | 0.720075758 | 0.719010385 |
| Pain | 600 | 0.714689266 | 0.718843553 | 0.708472187 | 0.708641975 |
| ADR Count | 100 | 0.574324324 | 0.587815338 | 0.574266226 | 0.577359036 |
| Disease Count | 400 | 0.564032698 | 0.587314081 | 0.56522314 | 0.572110622 |
| Mental Count | 100 | 0.707650273 | 0.711343831 | 0.706451613 | 0.705537802 |
| Coughing | 100 | 0.666191155 | 0.665535147 | 0.664103673 | 0.664272031 |

**Random Forests without SMOTENC:**

From the given charts, the following can be observed:

* The metrics have now become less equal, with accuracy increasing dramatically by the F-score giving much lower results.
* The previous observation for weight/height and pressure is now reversed, this indicates how the continuous values have a much greater influence in the SMOTENC algorithm.
* Count predictions is now extremely low.
* AskAPatient dataset’s performance have dropped massively.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Complete Subset | Best Estimator | Accuracy | Precision | Recall | FScore |
| Hypertensive disease | 500 | 0.598290598 | 0.599862732 | 0.595723684 | 0.592870232 |
| Pain | 200 | 0.709401709 | 0.552476669 | 0.536667335 | 0.53618795 |
| ADR Count | 300 | 0.423076923 | 0.39879021 | 0.388964322 | 0.391975191 |
| Disease Count | 500 | 0.514957265 | 0.41622575 | 0.377881411 | 0.377375272 |
| Mental Count | 500 | 0.542735043 | 0.35175836 | 0.357624831 | 0.352198359 |
| Weights Subset | Best Estimator | Accuracy | Precision | Recall | FScore |
| Hypertensive disease | 300 | 0.820512821 | 0.777777778 | 0.792207792 | 0.784189723 |
| Pain | 400 | 0.717948718 | 0.621428571 | 0.555194805 | 0.546031746 |
| ADR Count | 500 | 0.333333333 | 0.33015873 | 0.33015873 | 0.329885057 |
| Disease Count | 400 | 0.615384615 | 0.415343915 | 0.441414141 | 0.42037037 |
| Mental Count | 100 | 0.512820513 | 0.348484848 | 0.358674464 | 0.351984696 |
| Pressure Subset | Best Estimator | Accuracy | Precision | Recall | FScore |
| Hypertensive disease | 100 | 0.712230216 | 0.697411003 | 0.661803714 | 0.667464115 |
| Pain | 600 | 0.741007194 | 0.684752747 | 0.667371827 | 0.674048984 |
| ADR Count | 800 | 0.525179856 | 0.470063539 | 0.444813409 | 0.440220275 |
| Disease Count | 100 | 0.489208633 | 0.458802553 | 0.392770563 | 0.407304055 |
| Mental Count | 100 | 0.633093525 | 0.731884058 | 0.512784424 | 0.566431442 |
| Ask A Patient | Best Estimator | Accuracy | Precision | Recall | FScore |
| Hypertensive disease | 100 | 0.77092511 | 0.520792079 | 0.514935989 | 0.514318631 |
| Pain | 400 | 0.696035242 | 0.473707825 | 0.483678699 | 0.471291139 |
| ADR Count | 800 | 0.559471366 | 0.382986513 | 0.341900161 | 0.336908553 |
| Disease Count | 300 | 0.528634361 | 0.349698674 | 0.363578972 | 0.35632566 |
| Mental Count | 800 | 0.766519824 | 0.513917004 | 0.507078507 | 0.498143745 |
| Coughing | 300 | 0.59030837 | 0.560015898 | 0.555324866 | 0.554930112 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Complete Subset | Best Estimator | Accuracy | Precision | Recall | FScore |
| Hypertensive disease |  |  |  |  |  |
| Pain |  |  |  |  |  |
| ADR Count |  |  |  |  |  |
| Disease Count |  |  |  |  |  |
| Mental Count |  |  |  |  |  |
| Weights Subset | Best Estimator | Accuracy | Precision | Recall | FScore |
| Hypertensive disease |  |  |  |  |  |
| Pain |  |  |  |  |  |
| ADR Count |  |  |  |  |  |
| Disease Count |  |  |  |  |  |
| Mental Count |  |  |  |  |  |
| Pressure Subset | Best Estimator | Accuracy | Precision | Recall | FScore |
| Hypertensive disease |  |  |  |  |  |
| Pain |  |  |  |  |  |
| ADR Count |  |  |  |  |  |
| Disease Count |  |  |  |  |  |
| Mental Count |  |  |  |  |  |
| Ask A Patient | Best Estimator | Accuracy | Precision | Recall | FScore |
| Hypertensive disease |  |  |  |  |  |
| Pain |  |  |  |  |  |
| ADR Count |  |  |  |  |  |
| Disease Count |  |  |  |  |  |
| Mental Count |  |  |  |  |  |
| Coughing |  |  |  |  |  |