

## BAB V

### SIMPULAN DAN SARAN

#### 5.1 Simpulan

PT Kredibel Teknologi Indonesia membutuhkan model machine learning dalam melakukan klasifikasi adverse media dalam membantu memilih berita untuk sistem customer due diligence. Penelitian ini melakukan pembuatan model *machine learning* dengan menggunakan algoritma XGBoost dan Gradient Boosting untuk melakukan klasifikasi *adverse media* dan bukan *adverse media* pada media daring. Model ini menggunakan judul berita yang sudah diberikan label oleh *annotator* yaitu sebagai *adverse media* atau bukan *adverse media*. Proses labeling data dilakukan dengan menggunakan *bot* Telegram dan hasil dari labeling data akan disimpan di dalam *database bot*. *Bot* Telegram yang dibuat menggunakan bahasa pemrograman Node.JS dan menggunakan *database* MariaDB.

Keakuratan pada model yang dihasilkan dalam melakukan klasifikasi *adverse media* dan bukan *adverse media* menjadi ukuran dalam efektivitas model yang dihasilkan. Model klasifikasi yang dibuat menggunakan bahasa pemrograman Python dengan menerapkan algoritma XGBoost dan Gradient Boosting. Hasil dari model yang dibuat akan disisipkan pada *framework* Flask dan pengguna dapat menggunakan model ini melalui halaman website maupun melalui pemanggilan API. Model yang digunakan yaitu model Gradient Boosting dengan tingkat akurasi sebesar 82.31% pada data *test* dan 84.93% pada data *validation*.

Model dari penelitian ini dapat digunakan Kredibel untuk melakukan klasifikasi berita. Proses klasifikasi ini akan mempermudah Kredibel dalam mengelompokkan berita yang akan digunakan dalam sistem *customer due diligence*. Model ini sudah bisa digunakan dalam proses klasifikasi namun masih perlu dikembangkan lebih lanjut karena keterbatasan data yang digunakan.

Model yang dihasilkan yaitu model dari algoritma Gradient Boosting dengan nilai akurasi sebesar 82.31% pada data *test* dan 84.93% dari data *validation*. Meskipun model-model ini memiliki tingkat akurasi yang cukup, perlu diingat bahwa keterbatasan data yang digunakan dalam penelitian ini mempengaruhi performa model. Jumlah data yang digunakan hanya sebanyak 753, yang masih dianggap kurang. Oleh karena itu, untuk meningkatkan kinerja model ini, diperlukan lebih banyak data dan variasi kombinasi kata yang lebih kompleks.

Selain itu, model-model ini juga mungkin tidak dapat secara efektif menghadapi data baru yang belum pernah diajarkan pada saat pelatihan. Variasi judul berita yang beredar sangat beragam, dan model ini mungkin tidak dapat mengatasi semua variasi tersebut dengan tingkat akurasi yang tinggi. Dengan demikian, penelitian ini menyadari adanya keterbatasan dan menyarankan untuk menggunakan lebih banyak data dalam penelitian mendatang guna meningkatkan kinerja model.

## 5.2 Saran

Berdasarkan hasil dari penelitian ini, terdapat beberapa saran yang bisa digunakan dalam penelitian berikutnya mengenai klasifikasi *adverse media* dengan menggunakan model *machine learning* yaitu:

- a. Melakukan eksplorasi lebih lanjut mengenai *hyperparameter* yang pada algoritma XGBoost maupun Gradient Boosting untuk mendapatkan performa terbaik karena terdapat banyak *hyperparameter* yang tidak mungkin diuji pada penelitian ini.
- b. Melakukan seleksi dan memperbanyak jumlah data dan variasi yang digunakan untuk meningkatkan performa terutama pada data baru yang membuat peningkatan kesalahan prediksi.
- c. Menggunakan isi berita untuk menghasilkan tingkat prediksi yang lebih baik dibandingkan hanya menggunakan judul berita.
- d. Terdapat risiko adanya *overfitting* pada model klasifikasi teks terutama dalam melakukan klasifikasi dengan data baru. Karena itu perlu adanya

analisis mendalam seperti penentuan *threshold* dari tingkat selisih dari *precision* dan *recall* antara data *train* dengan data *test* dan *validation* untuk mengetahui tingkat overfitting pada model.

- e. Melakukan *monitoring* dan peningkatan pada model yang dibuat secara berkala sesuai dengan berita terbaru yang ada saat ini.



UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA