

ABSTRAK

Data dari Kementerian Pekerjaan Umum dan Perumahan Rakyat (Kementerian PUPR) tahun 2019, sekitar 81 juta generasi milenial tidak memiliki rumah. Peraturan Pemerintah Nomor 25 Tahun 2020 tentang Penyelenggaraan Tabungan Perumahan Rakyat atau biasa disebut PP 25 Tapera Tahun 2020 menjadi usaha pemerintah agar masyarakat Indonesia bisa memiliki rumah. Tapera adalah penyimpanan berjangka oleh pekerja untuk pembiayaan rumah, atau bisa dikembalikan lagi setelah waktunya berakhir. Saat peraturan ini ditetapkan, banyak tanggapan masyarakat terkait peraturan ini. Berdasarkan tanggapan masyarakat tersebut, bisa diterapkan penelitian untuk melihat sentimen masyarakat terkait peraturan ini. Metode yang digunakan untuk penelitian ini adalah *support vector machine* (SVM) dikarenakan SVM memiliki tingkat akurasi yang baik. Penerapan SVM selain membutuhkan data latih, juga membutuhkan label, dan pelabelan secara manual membutuhkan waktu yang banyak. Metode pelabelan yang digunakan adalah *lexicon-based*. Kombinasi ini dilakukan dengan menggunakan *lexicon-based* untuk menentukan nilai sentimen atau label data dan data hasil *lexicon* dijadikan sebagai data pelabelan untuk inputan *Support Vector Machine classifier*. Kombinasi ini menjadikan *lexicon-based* sebagai sarana untuk mentransfer pembelajaran kepada SVM dengan cara memudahkan proses pemberian label. Proses penelitian diawali dengan pengumpulan data penelitian dari media sosial twitter, selanjutnya proses preprocessing untuk mengubah data kasar dan tidak terstruktur menjadi data yang siap digunakan, lalu pelabelan data dengan *lexicon-based*, kemudian pembobotan dengan TF-IDF, pengolahan menggunakan SVM, dan proses evaluasi model kinerja algoritma dengan *confusion matrix*. Hasil penelitian menunjukkan kombinasi *lexicon-based* dan SVM berjalan dengan baik. *Lexicon-based* berhasil memberikan label terhadap 640 data *tweet* dan data hasil pelabelan digunakan oleh SVM. SVM berhasil mendapatkan nilai *accuracy* 83.08% untuk *kernel linear*. *Kernel linear* juga memiliki hasil terbaik untuk nilai *recall*, dan *f1-score* dengan masing-masing nilai sebesar 63.19%, 64.56%. Sedangkan untuk nilai *precision*, *kernel rbf* memiliki hasil terbaik dengan 75.29%.

Kata Kunci: Analisa Sentimen, Tapera, Support Vector Machine, Lexicon Based