

INTISARI

PENERAPAN *LEXICON-BASED* UNTUK ANALISIS SENTIMENT MASYARAKAT TERHADAP KASUS KEBOCORAN DATA DI INDONESIA

Oleh:

JUAN KRISNA TANUBRATA

1911010076

Cybercrime adalah industri triliunan yang mengintai pengguna internet. Berdasarkan data yang dikumpulkan, pada tahun 2019 – 2022 terjadi sebanyak 5 kasus kebocoran data yang sangat besar pada perusahaan multinasional dan juga instansi pemerintahan. Ditengah kondisi terjadinya kebocoran data saat ini tingkat kepercayaan masyarakat kepada Pemerintah menjadi berkurang secara signifikan. Berdasarkan data yang diambil dari situs CNN Indonesia sebanyak 46% masyarakat ragu dengan keamanan data pribadi mereka, 43% masyarakat masih percaya, 10% masyarakat agak percaya dan 1% masyarakat tidak percaya sama sekali. Hal ini memungkinkan untuk melakukan analisa dan penelitian lebih lanjut mengenai presentase tingkat kepercayaan masyarakat terhadap Pemerintah. Pada penelitian ini dilakukan pengumpulan data dengan menggunakan teknik *text mining* di mana teknik ini akan menggali informasi yang berkualitas dari suatu rangkaian teks yang terangkum dalam sebuah dokumen. Tujuan dari penelitian ini adalah untuk mengumpulkan, mengukur dan menganalisa data opini masyarakat umum yang dituangkan melalui media sosial tentang kinerja pemerintahan Indonesia atas kebocoran data. Penelitian ini menggunakan metode *lexicon-based approach* dengan pendekatan *dictionary-Based Approach*. Sebagai tahapan analisa *sentiment* yang dimulai dari pengumpulan data dengan jumlah 15.000 data text, membersihkan data, pemuatan kamus yang memiliki pembobotan dengan range nilai -5 sampai +5, *extract* data, dan *determine sentiment* menjadi 3 klasifikasi (positif, negative dan netral). Selain menggunakan kamus yang dimuat sebelumnya, penelitian ini juga membuat kamus baru yang berisi jenis dan tindakan serangan berbasis komputasi agar dalam penentuan nilai sentiment dalam penelitian ini lebih mengerucut. Hasil dari analisa tersebut mendapatkan 23.0% *sentiment positive*, 62.6% *sentiment negative*, dan 14.3% *sentiment neutral*. Angka ini juga dibandingkan dengan analisa ahli Bahasa yang menghasilkan tingkat keakuratan 50%.

Kata Kunci: *lexicon-based approach, Cybercrime, kebocoran data, dictionary-Based Approach, Analisa sentiment.*

ABSTRACT

IMPLEMENTATION OF LEXICON-BASED FOR ANALYSIS OF PUBLIC SENTIMENT ON DATA LEAK CASES IN INDONESIA

By:
Juan Krisna Tanubrata

Cybercrime is an industry of trillions that stalks internet users. Based on the data collected, in 2019 - 2022, there were 5 cases of massive data leaks in multinational companies and government agencies. Amid the current data leak, the Government's public trust level has significantly reduced. Based on data taken from the CNN Indonesia website, as many as 46% of people doubt the security of their personal data, 43% of people still trust it, 10% have some confidence and 1% don't trust it at all. Many people are doubtful about their data security makes it possible to conduct further analysis and research regarding the percentage level of public trust in the Government. In this study, data collection was carried out using text mining techniques which this technique will extract quality information from a series of texts summarized in a document. The purpose of this study is to collect, measure and analyze data on public opinion as expressed through social media about the performance of the Indonesian Government on data leakage. This study uses a lexicon-based approach with a dictionary-based approach. The sentiment analysis stage starts with collecting data with a total of 15,000 text data, cleaning the data, loading dictionaries with weights with a value range of -5 to +5, extracting data, and determining sentiment into three classifications (positive, negative and neutral). In addition to using the preloaded dictionary, this research also creates a new dictionary which contains the types and actions of computation-based attacks to narrow the determination of sentiment value in this research. The results of this analysis get 23.0% positive sentiment, 62.6% negative sentiment, and 14.3% neutral sentiment. This figure is also compared with the analysis of linguists, which produces an accuracy rate of 50%.

Keywords: Lexicon-Based Approach, Cybercrime, Data Leakage, Dictionary-Based Approach, Sentiment Analysis.