

# Implementasi sentiment analysis dan topic modelling terhadap aduan konsumen di twitter: Studi kasus PT. XYZ = Implementation of sentiment analysis and topic modelling for customer complaint detection on twitter: A case study of PT. XYZ.

Anindito Izdihardian Wibisono, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20516844&lokasi=lokal>

---

## Abstrak

Pada tahun 2020, nilai customer satisfaction index (CSI) PT XYZ yang mempresentasikan kepuasan konsumen XYZ berjumlah 83.9. Angka ini gagal mencapai target PT XYZ di tahun tersebut yaitu 87, dan turun dari tahun sebelumnya yaitu 86,5 di tahun 2019. Berdasarkan pengambilan data, diketahui bahwa XYZ mengelola aduan konsumen hanya melalui Twitter. Dari ribuan tweet yang diterima akun resmi customer care PT XYZ (@XYZCares) tiap bulan di Twitter, diperkirakan hanya 1-2% yang dideteksi sebagai aduan dengan proses pengawasan manual. Penelitian ini merancang solusi dua langkah berupa implementasi social media listening dalam bentuk sentiment analysis dan topic modelling, untuk mengetahui isu dalam tweet aduan kepada XYZ. Dataset berupa kumpulan tweet yang menyebutkan @XYZCares pada kurun waktu 1 Januari 2020 - 31 Desember 2020. Data di-scrape dari Twitter menggunakan script Python. Hasil evaluasi secara cross-validation menunjukkan akurasi rerata sentiment analysis dengan algoritme SVM lebih akurat (77%) untuk kasus ini dibandingkan algoritme RF (75%). Untuk task pemodelan topik, algoritme LDA menghasilkan klaster topik sejumlah 4 dengan rerata TPC sebesar 80%. Diketahui bahwa topik yang dominan adalah isu korupsi dan suap di badan PT XYZ. Dengan mempertimbangkan penemuan tersebut, saran yang dapat diberikan berdasarkan penelitian ini adalah memberhentikan staf yang diduga terlibat dalam isu-isu tersebut, serta menerapkan good corporate governance berupa aspek pengawasan dan pencegahan korupsi.

.....The customer satisfaction index (CSI) for the year 2020 is calculated at 83.9. This value fails to reach the company's target for the year at 87 and is lower than the CSI value for 2019 at 86.5. Data acquired from the company shows that consumer complaints are accepted and processed only through Twitter. It is estimated that of the thousands of tweets processed by PT XYZ's official customer care account (@XYZCares) each month, only 1-2% of the tweets are considered complaints based on manual searching and classification. This research proposes a two-step solution by implementing social media listening in the form of sentiment analysis and topic modelling, to detect the most frequent issues addressed to XYZ. The dataset consists of tweets created from January 1st, 2020, to December 31st, 2020 which mentioned @XYZCares. The tweets were scraped from Twitter using Python scripts. The results of cross-validation show that for the task of sentiment analysis, SVM is a more accurate algorithm on average (77%) compared to Random Forest (75%). For the following task of topic modelling, the LDA algorithm model produced 4 topic clusters with an average TPC of 80%. The most dominant topic detected relate to allegations of bribery and corruption within PT XYZ. Taking these finds into consideration, this research suggests that PT XYZ immediately dismiss all staff implicated in the aforementioned cases, as well as implementing good corporate governance in the form of tighter supervision and prevention of corrupt dealings.