

BAB 2

LANDASAN TEORI

2.1 User Feedback

Feedback adalah suatu reaksi yang diberikan oleh komunikan dimana reaksi tersebut bisa menjadi reaksi positif atau negatif (Sunarjo 1983). *Product reviews* memiliki informasi yang penting mengenai masalah pelanggan dan pengalaman mereka dengan produk. Informasi tersebut dianggap penting untuk kecerdasan bisnis perusahaan yang dapat digunakan untuk tujuan konseptual desain, personalisasi, rekomendasi produk, pemahaman pelanggan yang lebih baik, dan menarik lebih banyak pelanggan (Zhan dkk., 2008).

2.2 TF-IDF

Metode TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada *information retrieval*. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. Metode ini akan menghitung bobot setiap token t di dokumen d dengan rumus pada persamaan (2.1) (Maarif, 2015).

$$w_{dt} = tf_{dt} \times \log\left(\frac{N}{dt}\right) \quad (2.1)$$

Keterangan:

d = dokumen ke- d

t = kata ke- t dari kata kunci

w = bobot dokumen ke- d terhadap kata ke- t

tf = banyaknya kata yang dicari pada sebuah dokumen
 N = total dokumen
 dt = banyak dokumen y

2.3 Latent Semantic Indexing

LSI (Latent Semantic Indexing) adalah metode pengindeksan otomatis yang memproyeksikan *documents* dan *terms* ke dalam ruang berdimensi rendah yang merepresentasikan konsep *semantic* dalam *document*. LSI didasarkan pada SVD (Singular Value Decomposition). SVD adalah metode yang digunakan untuk mendekomposisi matrix menjadi beberapa komponen matriks (Aggarwal dkk., 2012). SVD menggunakan rumus pada persamaan (2.2).

$$A = U S V^T \quad (2.2)$$

Keterangan:

U = m x m orthogonal matrix

S = m x n diagonal matrix dengan nilai non-negatif

V^T = n x n transpose orthogonal matrix

2.4 Support Vector Machine

Support Vector Machine adalah algoritma pembelajaran mesin yang dapat digunakan untuk masalah klasifikasi atau regresi. SVM menggunakan teknik yang disebut *kernel* untuk mengubah data dan kemudian berdasarkan transformasi ini, SVM menemukan batas optimal antara kemungkinan output (Aggarwal dkk., 2012). *Kernel* yang dipakai dalam penelitian ini adalah RBF. Sebelum melakukan proses klasifikasi menggunakan SVM, data akan dibagi menjadi dua kelas yaitu

kelas 1 (puas) dan kelas -1 (tidak puas). Support Vector Machine menggunakan rumus pada persamaan (2.3).

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (2.3)$$

Keterangan:

x_i = data ke -i

w = nilai bobot vektor

b = nilai bias

y_i = target data ke -i

ξ_i = variabel slack untuk data ke-i

perhitungan nilai b menggunakan persamaan (2.4) dan persamaan (2.5) untuk menghitung nilai w .

$$b = y_i - w^T x_i \quad (2.4)$$

$$w = \sum_{i=1}^n \alpha x_i y_i \quad (2.5)$$

Keterangan:

b = nilai bias

w = bobot vektor

α = lagrange multiplier, nilai lagrange = 0.1

x_i = data ke -i

y_i = target data ke -i

Perhitungan optimal *hyperplane* menggunakan persamaan (2.6).

$$\min \frac{1}{2} |w|^2 + C(\sum_{i=1}^n \xi_i) \quad (2.6)$$

Keterangan:

w = bobot vektor

C = parameter yang menentukan besar penalti

ξ_i = variabel slack untuk data ke-i

Bentuk primal svm menggunakan rumus pada persamaan (2.7).

$$L_p(w, \xi, b, \alpha) = \frac{1}{2} |w|^2 + C \left(\sum_{i=1}^n \xi_i \right) - \sum_{i=1}^n \alpha_i \{y_i(x_i \cdot w + b) - 1 + \xi_i\} - \sum_{i=1}^n \mu_i \xi_i \quad (2.7)$$

Keterangan:

b = nilai bias

w = bobot vektor

C = parameter yang menentukan besar penalty

ξ_i = variabel slack untuk data ke-i

x_i = data ke -i

y_i = target data ke -i

α = lagrange multiplier

μ = lagrange multiplier

Kernel RBF menggunakan rumus pada persamaan (2.8).

$$K(X_1, X_2) = \exp\left(-\frac{\|X_1 - X_2\|^2}{2\sigma^2}\right) \quad (2.8)$$

Keterangan:

σ = nilai varian

$\|X_1 - X_2\|$ = jarak euclidean antara poin X_1 dan X_2

2.5 Confusion Matrix

Confusion matrix adalah metode untuk mengukur seberapa baik performa model Support Vector Machine dalam melakukan klasifikasi data. Confusion Matrix adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada confusion matrix yaitu (True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN)).

Tabel 2.1 Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True positive (TP)	False positive (FP)
Predicted Negative (0)	False negative (FN)	True negative (TN)

Setelah menghitung (TP, TN, FP, FN) maka dapat menghitung accuracy, precision, recall dan F1-score.

$$Accuracy = \frac{TP}{TP + FP} \quad (2.9)$$

$$Precision = \frac{TP}{TP + FP} \quad (2.10)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.11)$$

$$F1 = \frac{(2 * Recall * Precision)}{TP + FN} \quad (2.12)$$

Keterangan:

TP = Jumlah data dengan nilai sebenarnya positif, dan nilai prediksi positif

TN = Jumlah data dengan nilai sebenarnya negatif, dan nilai prediksi negatif

FP = jumlah data dengan nilai sebenarnya negatif dan nilai prediksi positif.

FN = jumlah data dengan nilai sebenarnya negatif dan nilai prediksi negatif.