

**PROGRAM STUDI SARJANA SISTEM INFORMASI
PROPOSAL PROYEK PEMROSESAN BAHASA ALAMI**



***Analysis Emotion Of Comment Topic Reddit With Multi-label
Classification Using SVM Algorithm***

Oleh:

12S18027 Christopher A Hutabarat

12S18028 Alda G M Lumban Gaol

12S18032 Sarah H M Siahaan

**FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO
INSTITUT TEKNOLOGI DEL
LAGUBOTI
NOVEMBER 2021**

DAFTAR ISI

DAFTAR TABEL.....	3
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Tujuan.....	2
1.3 Manfaat.....	2
1.4 Ruang Lingkup	2
BAB 2 RENCANA KERJA.....	3
2.1 Pembagian Tugas.....	3
2.2 Jadwal Kerja	3
REFERENSI	4

DAFTAR TABEL

Tabel 1 Pembagian Tugas	3
Tabel 2 Jadwal Kerja.....	3

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Reddit (reddit.com) adalah sebuah situs yang dibangun oleh Alexis Ohanian dan Steve Huffman pada tahun 2005. Reddit merupakan sebuah situs penyedia konten, dimana para penggunanya yang disebut sebagai *redditors* dapat membagikan maupun memberikan komentar terhadap konten-konten yang ada pada reddit. Konten dan komentar tersebut dapat disukai (*upvote*) atau tidak disukai (*downvote*) oleh *redditors* lain. Konten dan komentar *redditors* yang mendapat *downvote* akan diberikan poin karma, dan yang mendapat *upvote* akan mendapatkan penghargaan. Hal ini menjadi pendorong bagi para *redditors* untuk memposting konten yang baik, membuat komentar yang bermanfaat dan juga memberikan umpan balik yang relevan [1].

Salah satu cara yang dapat dilakukan untuk menilai kualitas komentar yang *diposting* pada reddit adalah dengan menggunakan *emotion analysis*. *Emotion analysis* adalah proses mengidentifikasi dan menganalisis emosi yang diungkapkan dalam suatu data tekstual. *Emotion analysis* dalam teks dapat membantu menentukan opini dan maksud afektif dari penulis, serta sikap, evaluasi, dan kecenderungan mereka terhadap berbagai topik [2].

Memahami opini publik dapat membantu kreator dalam mengambil keputusan [3]. Sama halnya dengan *redditors* pada *platform* reddit yang sebaiknya juga memahami opini *redditors* lainnya terhadap konten yang dihasilkan dan komentar yang diberikan. Sehingga peluang *redditors* akan semakin besar dalam menghasilkan konten dan memberikan tanggapan yang ditanggapi dengan aksi *upvote*.

Multilabel classification merupakan sebuah *task* prediktif pada *data mining* yang memungkinkan untuk mengasosiasikan sebuah data kedalam beberapa label. Setiap konten dan komentar yang ada pada reddit akan diberikan label dan bisa saja setiap konten dan komentar memiliki label lebih dari satu. Adapun label pada *dataset* tersebut digunakan untuk mempermudah proses klasifikasi. Terdapat beberapa metode klasifikasi yang dapat digunakan dalam klasifikasi multilabel data, namun pada proyek ini akan digunakan metode klasifikasi *Support Vector Machine* (SVM). Pada dasarnya SVM digunakan untuk menentukan *decision boundary* atau *hyperplane* antar dua kelas, sehingga akan menjadi tantangan apabila diimplementasikan pada *multilabel classification* karena secara *native* cara kerja pengklasifikasian SVM berbeda dengan klasifikasi multilabel. Sehingga, untuk mengatasi tantangan tersebut dapat digunakan *MultiOutputClassifier* SVM yang disediakan oleh *machine learning open source library* berbasis *python* yaitu *Scikit-*

learn. Strategi ini merupakan pengklasifikasian multi target yang dapat digunakan untuk mendukung klasifikasi multilabel dengan SVM.

Oleh karena itu, tim proyek memutuskan untuk mengangkat proyek dengan judul “*Analysis Emotion Of Comment Topic Reddit With Multi-label Classification Using SVM Algorithm*”. Proyek diharapkan dapat berkontribusi memecahkan tantangan multilabel *classification* menggunakan SVM, menganalisis emosi pada komentar reddit dengan *multilabel classification* menggunakan metode SVM, dan melakukan pengujian akurasi terhadap metode tersebut.

1.2 Tujuan

Tujuan dari pelaksanaan proyek ini adalah:

1. Berkontribusi memecahkan tantangan *multilabel classification* menggunakan metode SVM.
2. Menerapkan metode SVM dalam menganalisis emosi pada komentar di reddit dengan *multilabel classification*.
3. Untuk mengetahui bagaimana tingkat akurasi menggunakan metode SVM dalam melakukan *emotion analysis*.

1.3 Manfaat

Manfaat dari pelaksanaan proyek ini adalah:

1. Bagi *redditors*
Proyek ini diharapkan dapat menjadi bahan evaluasi bagi para *redditors* dalam membagikan konten yang lebih baik.
2. Bagi Mahasiswa
Proyek ini diharapkan dapat memberi wawasan bagi tim proyek dalam menerapkan pemrosesan bahasa alami tepatnya *SVM algorithm* pada data teks reddit.

1.4 Ruang Lingkup

Ruang lingkup dari pelaksanaan proyek ini adalah:

1. Dataset yang digunakan adalah dataset *g0_emotions* yang berasal dari Hugging Face yang berisi 58000 komentar reddit dengan label sebanyak 27 kategori.
2. Metode yang digunakan dalam pengerjaan proyek *multilabel classification* adalah *Support Vector Machine* (SVM).

BAB 2 RENCANA KERJA

2.1 Pembagian Tugas

Tabel 1 Pembagian Tugas

NIM	NAMA	KONTRIBUSI PEMBAGIAN TUGAS
12S18027	Christopher A Hutabarat	33,33%
12S18028	Alda G M Lumban Gaol	33,33%
12S18032	Sarah H M Siahaan	33,33%

Berikut tugas yang dilakukan oleh tim proyek:

- **Data Analyst**

Mengidentifikasi dan menafsirkan data model kemudian melakukan analisis terhadap strategi yang efisien untuk diimplementasikan pada proyek.

- **Programmer**

Merancang dan mengimplementasikan *code* untuk membangun sistem sesuai hasil analisis *data analyst*.

2.2 Jadwal Kerja

Tabel 2 Jadwal Kerja

No	Task	Minggu ke-			
		1	2	3	4
1.	Membuat Proposal Proyek - Analisis Latar Belakang - Pemilihan Metode - Presentasi Proposal				
2.	Desain - Data Preprocessing - Feature Extraction - Feature Selection - Modelling with SVM				
3.	Implementasi - Exploratory Data Analysis (EDA) - Data Preprocessing - Feature Extraction - Feature Selection - Modelling with SVM				
4.	Evaluasi Akurasi Model				
5.	Laporan Akhir dan Presentasi Akhir				

REFERENSI

- [1] Anderson, "Ask me anything: what is Reddit?," *Library Hi Tech News*, vol. 32, pp. 8-11, 2015.
- [2] S. Aman and S. S. , "Identifying expressions of emotion in text," in *International Conference on Text, Speech and Dialogue*, Springer, Berlin, Heidelberg, 2007.
- [3] J. K. Rout, "A model for sentiment and emotion analysis of unstructured social media text," *Electron Commer Res* , no. 18, pp. 181-199, 2017.
- [4] K. I. Gunawan and J. Santoso, "Multilabel Text Classification Menggunakan SVM dan Doc2Vec Classification pada Dokumen Berita Bahasa Indonesia," *Journal of Information System, Graphics, Hospitality and Technology*, vol. 3, no. 1, pp. 29-38, 2021.