

1. Pendahuluan

1.1 Latar Belakang

Pada saat ini, media sosial sudah tidak dapat dipisahkan dengan kehidupan manusia. Media sosial menyediakan platform yang nyaman untuk melakukan komunikasi sosial [1]. Hal tersebut membuat pengguna akan sangat cepat mendapatkan informasi yang tersebar sangat luas di internet. Hal ini membuat semua informasi yang ada di internet menjadi dipertanyakan kebenaran dan sumbernya. Media sosial menyimpan dan menyebarkan opini pribadi, fakta, berita bohong, dan berbagai informasi yang mungkin disukai oleh pengguna. Oleh karena itu semakin sering sebuah topik dibicarakan pada media sosial, semakin banyak juga argumen – argumen yang diberikan oleh para pengguna pada suatu topik tersebut. Contohnya seperti kasus pinjaman *online* yang akhir – akhir ini sedang ramai dibicarakan. Hal ini menyebabkan banyaknya argumen mengenai kasus pinjaman *online* ilegal yang dimana masing – masing pengguna dapat menyampaikan argumennya mengenai pinjaman *online* ilegal tersebut.

Natural Language Processing (NLP) adalah suatu kumpulan dari teknik komputasi yang dimana memiliki fungsi untuk melakukan analisis dan representasi dari Bahasa manusia [2]. NLP adalah dapat diklasifikasi menjadi dua bagian, yaitu *Natural Language Understanding* (NLU) dan *Natural Language Generation* (NLG). NLU dapat membantu mesin untuk memahami *natural language* dan melakukan analisis dengan cara mengekstrak dari konsep teks, emosi, *keyword*, dan lainnya. Beberapa level dalam NLP yaitu seperti *Phonology*, *Syntax*, *Morphology*, *Semantics*, dan *Pragmatics* [3]. *Machine learning* (ML) adalah pembelajaran komputer yang dimana konsep pada komputer sains yang, statistika, ilmu kognitif, optimasi teori, dan banyak disiplin ilmu lainnya pada matematika dan sains. *Machine Learning* dapat diklasifikasikan menjadi dua bagian yaitu *supervised machine learning* dan *unsupervised machine learning*. *Supervised learning* adalah teknik untuk mencari hubungan antara atribut *input* (variabel yang independen) dengan atribut target (variabel yang sudah ditentukan). *Supervised* dapat diklasifikasikan menjadi dua kategori yaitu klasifikasi dan regresi. Klasifikasi adalah pendekatan pada *data mining* (*machine learning*) yang digunakan untuk prediksi pada suatu data [4].

Dari beberapa penelitian yang sudah dilakukan sebelumnya, banyak yang sudah melakukan *argument mining* dengan menggunakan data dari media sosial *Twitter*. Seperti pada tahun 2017, M. Dusmanu., dkk [5] melakukan pengembangan terhadap sistem *argument mining* pada media sosial *Twitter* dengan mengklasifikasikan berdasarkan fakta, opini, dan sumber berita. Metode yang digunakan adalah metode klasifikasi *Logistic Regression* (LR) dan *Random Forest* (RF). Hal yang dilakukan dilakukan dalam pengembangan ini adalah melakukan pemisahan *Tweets* yang merupakan argumen dengan non-argumen, lalu memisahkan *tweets* yang mengandung *tweets* yang factual atau hanya opini, dan terakhir melakukan pengecekan apakah *tweets* yang factual tersebut dilengkapi dengan sumber berita atau tidak. Hasil perhitungan yang didapatkan adalah dengan nilai *Precision* 69 %, lalu *Recall* 64 %, dan *F1-Score* 67 % . Selain itu terdapat juga penelitian yang dilakukan oleh Bosc, T. [6] Pada tahun 2016 melakukan *argument mining* yang dimana mengembangkan metodologi yang bernama DART (*Dataset of Arguments and Their Relations on Twitter*). Metode yang dikembangkan ini adalah untuk mencari hubungan antara argumen dengan argument yang lainnya. Data yang digunakan pada penelitian ini adalah topik mengenai politik mengenai Iran, lalu politik mengenai *Grexit*, lalu peluncuran produk dari *Apple iWatch*, dan terakhir penayangan episode 4 dari serial televisi *Game of Thrones*. Tahapan yang dilakukan adalah dengan melakukan *arguments annotation*, yaitu proses pemberian label pada sepasang *tweets*. Lalu tahap selanjutnya yaitu membuat pasangan *tweets* dengan topik yang sejenis. Tahapan akhir adalah dilakukan *argument linkin* yang membuat *tweets* A dan *Tweets* B termasuk ke dalam *Support*, *Attack*, atau *Unknown* .

Algoritma klasifikasi yang digunakan oleh penulis pada penelitian ini adalah k-Nearest Neighbor atau k-NN. k-Nearest Neighbor adalah metode klasifikasi yang paling dasar dan paling sederhana dan akan menjadi pilihan pertama untuk melakukan studi klasifikasi ketika tidak terlalu banyak persebaran data [7]. Penulis menggunakan metode k-NN pada penelitian ini dikarenakan metode klasifikasi k-NN adalah metode klasifikasi yang tidak memerlukan parameter yang pasti, melainkan berdasarkan dari besarnya data latih. k-NN juga dapat menghasilkan performansi yang baik untuk data yang memiliki persebaran data yang tidak merata. Dan juga metode klasifikasi k-NN memiliki beberapa cara perhitungan jarak seperti *euclidean distance*, *manhattan distance*, dan *minkowski distance*. Hal itu yang membuat pengujian dapat dilakukan dengan kombinasi dari parameter yang cukup beragam. Lalu, k-NN juga dapat melakukan permasalahan klasifikasi dan juga regresi dalam masalah memprediksi masalah. Tetapi metode ini lebih ringkas dan baik untuk digunakan dalam melakukan dan mengeksekusi pada permasalahan klasifikasi. Maka dari itu, tugas akhir ini akan menganalisis klasifikasi *tweets* termasuk ke dalam argumen atau non argumen pada media sosial *Twitter* yang diterapkan untuk *tweets* Bahasa Indonesia, yang berfokus pada proses klasifikasi berita mengenai pinjaman *online* ilegal pada *tweets* argumen atau *tweets* non-argumen.

1.2 Topik dan Batasannya

Berdasarkan dari uraian pada latar belakang, maka permasalahan dalam tugas akhir ini adalah diperlukannya sebuah sistem yang dapat melakukan klasifikasi *tweets* berdasarkan argumen atau non argumen. Proses klasifikasi *tweets* ini menggunakan metode klasifikasi k-*Nearest Neighbors*. Batasan masalah pada penulisan ini yaitu sebagai berikut:

1. Berfokus hanya pada platform media sosial *Twitter* dalam melakukan pencarian data dengan menggunakan *library Twint*.
2. Berfokus hanya pada tweets dengan Bahasa Indonesia dengan pembahasan sekitar pinjaman *online* ilegal (pinjol ilegal) pada tahun 2021.
3. Proses klasifikasi hanya menggunakan metode *k-Nearest Neighbor* dengan menggunakan *feature extraction* TF-IDF dan *Bag of Words*.

1.3 Tujuan

Berdasarkan dari uraian pada latar belakang, maka didapatkan tujuan dari pelaksanaan tugas akhir ini adalah sebagai berikut:

1. Membuat sistem klasifikasi *tweets* argumen dan non-argumen pada *Twitter*.
2. Menggunakan algoritma klasifikasi *k-Nearest Neighbors*.
3. Mendapatkan hasil evaluasi dari klasifikasi *k-Nearest Neighbors*.

1.4 Organisasi Tulisan

Organisasi pada penelitian ini pada bagian awal menjelaskan mengenai pendahuluan, bagian kedua menjelaskan studi terkait, bagian ketiga menjelaskan perancangan sistem yang dibangun, bagian keempat menjelaskan evaluasi pengujian dan bagian kelima menjelaskan kesimpulan.