

1. Pendahuluan

Latar Belakang

Text mining adalah proses ekstraksi pola berupa informasi dan pengetahuan yang berguna dari sejumlah besar sumber data teks, seperti dokumen word, PDF, kutipan teks, dll. Jenis masukan untuk penambangan teks ini disebut data tak terstruktur dan merupakan pembeda utama dengan penambangan data yang menggunakan data terstruktur atau basis data sebagai masukan.[1] Upaya pencarian dan penambangan data yang ada di dalam sebuah dokumen, bertujuan untuk mencari kata-kata yang dapat mewakili isi dari sebuah dokumen sehingga dapat dilakukan analisa keterhubungan pada setiap dokumen. *Text mining* memiliki banyak teknik salah satunya *Argumentation Mining*.

Argumentasi adalah sebuah alasan yang biasa digunakan untuk memperkuat atau membantah suatu pendapat, gagasan, atau pendirian.[2] Argumentasi biasanya dilakukan dengan lisan ataupun tulisan. *Argumentation mining* merupakan sebuah teknik yang berfokus pada ekstraksi dan analisis sebuah kalimat argumen dalam suatu teks bahasa. [2]. Argumentasi merupakan aktivitas yang bertujuan untuk meningkatkan atau mengurangi penerimaan sebuah sudut pandang bagi para pendengar atau pembaca, yang bertujuan untuk membenarkan sebuah sudut pandang yang lebih masuk akal [3]. Penggunaan *Argumentation Mining* ini dapat memudahkan penulis atau pembuat pernyataan dalam memeriksa teks yang lebih rasional dan meningkatkan kualitas sebuah argumentasi[4]. *Argumentation Mining* sendiri berfokus pada identifikasi kalimat argumen yang berupa komponen dan relasi antar kalimat argumen[5]. Pada komponen sebuah kalimat argumen, terdiri dari *claim* dan *premise*. *Claim* adalah komponen utama pada sebuah kalimat argumen yang dapat diterima atau ditolak pendengar atau pembaca. Sedangkan *premise* bertujuan untuk memvalidasi komponen *claim* agar pendengar atau pembaca dapat memastikan penerimaan pada *claim* tersebut tersebut atau malah sebaliknya.[6]. Identifikasi dan klasifikasi yang akan dilakukan menggunakan ekstraksi fitur, ekstraksi fitur adalah salah satu cara untuk melakukan identifikasi dan klasifikasi. Penggunaannya ekstraksi fitur mengambil ciri-ciri yang dapat menggambarkan karakteristik pada sebuah teks[7]

Permasalahannya, banyak orang yang menyebutkan teks argumen pada sebuah teks seperti teks artikel, teks esai, teks debat, teks novel dan lain lain. Isi dari teks tersebut memiliki kalimat pernyataan yang disebut *claim*. Tetapi tidak semua kalimat *claim* memiliki tujuan yang sama, seperti menunjukan sebuah kalimat argumentasi pada teks tersebut. Maka teks tersebut belum bisa dikatakan sebagai teks argument. Sehingga Pada penelitian ini, akan menerapkan klasifikasi argumen dengan *argumentation mining* pada data teks.. Untuk mencari tingkat akurasi suatu fitur pada kalimat argument. Dalam melakukan klasifikasi argument akan menggunakan *Argumentation mining*, dan akan diekstraksi menggunakan *lexical features* dan *contextual features*, karena *lexical features* dan *contextual features* masing-masing menggunakan cara ekstraksi yang berbeda. Setelah itu hasil klasifikasi dari *lexical features* dan *contextual features* akan diproses menggunakan metode *naïve bayes*. alasan penggunaan *lexical features* dan *contextual features* dikarenakan 2 fitur tersebut memiliki kesamaan pada saat melakukan pencarian pada kalimat. 2 fitur tersebut terfokus pada label kata-kata pada kalimat, yang dilabelkan pada saat *post tagging*.

Batasan Masalah

Pada penelitian cara mengklasifikasikan komponen argumen dari esai persuasif berbahasa Inggris yang berasal dari *website TU-Darmstadt* dengan hanya menggunakan ekstraksi fitur *lexical* dan *contextual*, model *naive bayes* sebagai klasifikasi. Dimana data teks sudah dilabelkan ke dalam bentuk *claim* dan *premise*. Sedangkan untuk nilai atribut fitur pada klasifikasi *premise* dan *claim* belum ada parameter yang pasti, sehingga nilai parameter berdasarkan dataset yang telah disediakan. Dan pada saat melakukan preprocessing menggunakan NLTK pada python.

Tujuan

Tujuan dari penelitian ini adalah menghitung hasil performansi dari ekstraksi fitur *lexical* dan *contextual* dalam mengklasifikasikan komponen argumen dengan menggunakan metode *naive bayes*. Kemudian evaluasi menggunakan 10-fold Validation untuk membagi data train dan data test lalu model evaluasi dihitung menggunakan nilai *accuracy*, *precision*, *recall*, dan F1-Score.