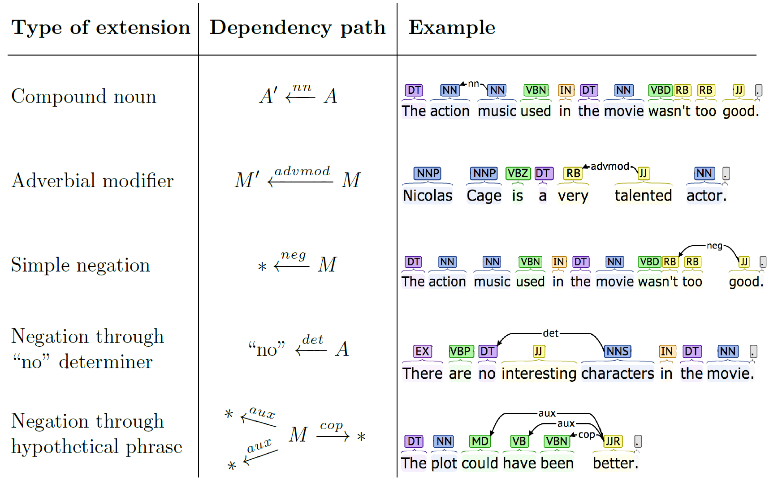
**LAPORAN *ASPECT EXTRACTION* DENGAN MENGGUNAKAN METODE *PATTERN BASED***

|  |  |
| --- | --- |
| Joshua Panjaitan  Telkom University  School of Computing  Bandung, Indonesiajoshuapanjaitan@student.telkomuniversity.ac.id | M Rakha Lutfi  Telkom University  School of Computing  Bandung, Indonesiamrakhalf@student.telkomuniversity.ac.id |

***Abstract*— Perkembangan teknologi saat ini begitu cepat, terutama dalam pertumbuhan jumlah data. Peningkatan penggunaan akun social media dan transaksi online masih menjadi komoditas panas dalam perkembangan teknologi saat ini. Dalam bidang traksaksi online, ulasan produk merupakan data yang tidak bisa ditinggalkan begitu saja, perlu dilakukan pengolahan untuk dapat menganalisis pasar didalam perkembangan teknologi saat ini. Sentimen analisis merupakan cabang ilmu yang mengolah data teks dan memberikan analisis terhadap data tersebut. Dalam sentimen analis terdapat tahapan ekstraksi aspek dan klasifikasi sentiment yang dimana tujuannya untuk menentukan aspek dan opini yang melekat pada suatu kalimat, yang nantinya akan digunakan dalam proses klasifikasi sentimen. Terdapat banyak metode untuk menemukan pasangan aspek dan opini, salah satunya adalah metode *rule based*, metode ini terbilang sederhana dan cenderung memberikan hasil yang tidak mengecewakan. Sedangkan dalam proses klasifikasi metode yang memanfaatkan korpus dalam ukuran bersar dapat dilakukan untuk menghemat proses komputasi. Pada penelitian ini akan dilakukan *aspect extraction* dengan metode *rule based* untuk menentukan pasangan aspek dan opini, dan menggunakan sentiwordnet untuk membantu proses klasifikasi sentimen kedalam positif dan negatif maupun netral. kemudian dilakukan pengujian untuk menentukan hasil akurasi dari metode tersebut.**

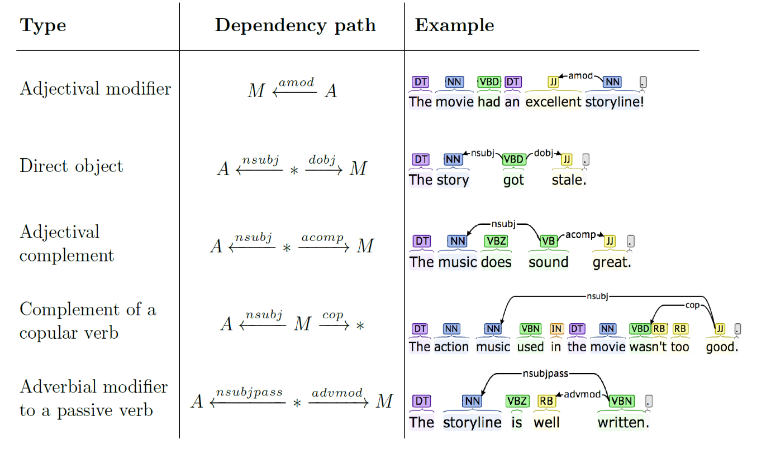
Keywords—aspect extraction, analysis, rule based, aspect.

# Introduction

*Aspect Extraction* merupakan salah satu tahapan dalam Sentimen Analisis yang dimana tujuannya adalah untuk mendapatkan pasangan aspek dan opini yang terdapat dalam sebuah kalimat. Pasangan aspek dan opini yang telah di ekstrak nantinya akan sangat berguna untuk tahapan selanjutnya dalam analisis sentimen, sehingga perlu untuk menemukan pasangan aspek dan opini yang sesuai untuk proses selanjutnya. Dalam menentukan aspek dalam sebuah kalimat terdapat beberapa cara yang bisa digunakan, salah satunya adalah metode *Pattern Rule Based.* Metode berbasis aturan yang telah didefinisikan sebelumnya oleh peneliti sebelumnya untuk menemukan pasangan aspek opini. Metode ini relatif mudah untuk diimplementasikan namun memiliki beberapa kekurangan dalam menemukan aspek yang sesuai karena hanya terbatas dalam aturan yang relatif kaku. Dalam penelitian ini metode *pattern based* dipilih untuk diimplementasikan dikarenakan kemudahan dalam implementasi dalam kasus yang tidak terlalu kompleks, dengan dataset ulasan camera Canon G3.

*Sentiment Classfication* Merupakan tahapan yang dilakukan setelah aspek dan opini sudah dipisahkan dari kalimat, tujuan dari tahapan ini adalah untuk memishkan sentiment penjadi 3 kategori yaitu positif dan negative maupun netral. Pada tahapn ini terdapat beberapa metode yang bisa digunakan seperti pendekatan dengan menggunakan machine learning ataupun menggunakan daftar sentimen lexicon yang berisi kata dan polaritas dari kata tersebut.

# PROPOSED METHOD

Dalam pengantar di atas telah dijelaskan bahwa metode yang digunakan dalam penelitian ini adalah metode *pattern rule based*. Metode ini memanfaatkan hubungan dari tiap-tiap kata untuk menemukan pasangan aspek dak opini. Hubungan dari tiap kata atau yang lebih sering disebut *dependency rule,* merupakan fitur utama dalam metode ini. *Dependency rule* bisa didapatkan dengan menggunakan library yang sudah ada, sehingga yang dibutuhkan hanyalah text input dan sistem akan otimatis mengeluarkan *dependency rule* dari tiap-tiap kata. *Rule* yang digunakan dalam *study case* ini menggunakan *rule* yang telah didefinisikan oleh Wouter Becken d.k.k didalam papernya yang berjudul Automatically Detecting and Rating Product Aspects from Textual Customer Reviews [1]. Wouter Becken di dalam papernya menggunakan 10 aturan untuk menentukan pasangan aspek dan opini seperti dalam gambar di bawah ini :

Sedangkan untuk proses klasifikasi akan menggunakan menanfaatkan sentiwordnet. Setiap sentimen yang sudah berhasil dipishkan akan dicocokan dengan sentiwordnet untuk mendapatkan nilai positive,negative dan objective dari sentiment tersebut. Setelah nilai tersebut diperoleh maka akan dapat diketahui polaritas dari sentimen tersebut.

# IMPLEMENTATION

## Aspect Extraction

Implementasi dilakukan dengan menggunakan bahasa pemrograman python dengan menggunakan beberapa library pemrosesan bahasa alami seperti nltk dan *stanford dependency parser*.Tahapan yang pertama yang dilakukan adalah proses membersihkan data ulasan, yang dimana data ulasam masih memiliki beberapa simbol dari hasil anotasi seperti ## yang sebelumnya di ikuti oleh aspek yang melekat dalam datanya seperti dalam contoh : “canon powershot g3[+3]##i recently purchased the canon powershot g3 and am extremely satisfied with the purchase “ , dimana dalam proses ini simbol anotasi akan dihilangkan dan aspek yang melekat akan disimpan untuk proses validasi/akurasi. Proses ini dilakukan dalam file clean.py. Output yang dihasilan dalam proses ini adalah sebuah dataset yang bersih berisi 600 baris yang setiap barisnya berisi ulasan yang siap untuk di ekstrak aspek dan opininya.

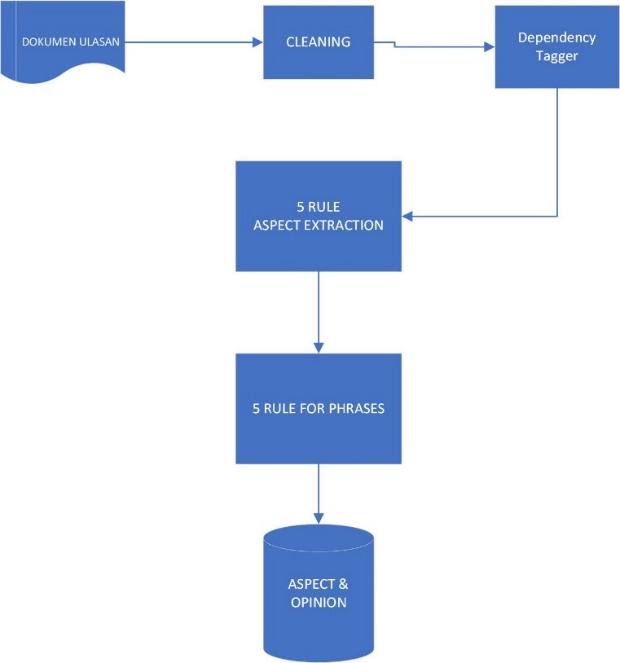


Fig 3 Gambaran Umum Sistem

Tahapan kedua adalah proses *tagger*. Di mana proses bertujuan untuk melabeli setiap kata dengan kata lain berdasarkan hubungan yang dibentuk dari tiap kata dalam kalimat. Seperti dalam gambar 1, kata Action Music memliki relasi NN, relasi dan hubungan kata ini merupakan tujuan utama yang akan dicari dalam proses ini. Proses ini menggunakan Dependency Tag dari library *Stanford Dependency Parser.* Semua keterkatian antar kata akan disimpan untuk proses selanjutnya.

Tahapan ketiga adalah proses implementasi *rule* yang telah didefinisikan dalam paper. 5 Rule pertama bertujuan untuk mencari pasangan aspek dan opini sedangkan 5 Rule berikutnya untuk mencari frasa dalam sebuah aspek maupun opini. Dalam tahapan ini setiap ulasan akan dicek pada setiap 5 rule aspek dan opini, jika kata ulasan masuk kedalam rule tersebut maka tahap selanjutnya memeriksa apakah aspek atau opini tersebut memiliki frasa, sehingga setiap aspek maupun opini bisa lebih dari 1 kata. Setelah semuanya didapat, hasil disimpan dalam list yang berisi pasangan aspek dan opini, di mana hasil ini akan disimpan dalam file .csv untuk proses perhitungan akurasi.

## Sentiment Classfication

Dalam proses ini akan menggunakan *library* NLTK untuk mendapatkan sentiwordnet. Setiap sentimen dari tiap kalimat akan dimasukkan diperiksa kedalam sentiwordnet untuk mendapatkan nilai polaritasnya, sehingga jika sentimen tersebut mendapatkan nilai polaritas maka hasil polaritas tersebut akan dilekatkan kedalam aspek yang melekat pada sentimen tersebut. Hasil dari implementasi ini adalah sebuah aspek dengan nilai sentien dan polaritas dari aspek tersebut,

Seperti ”Good Camera”, maka hasil dari klasifikasi sentimen ini adalah Camera [+].

# EVALUASI

## Aspect Extraction

Hasil yang didapat dalam percobaan ini disimpan dalam file Aspect Extraction.csv, di mana hasil yang didapat terlihat tidak begitu meyakinkan, dan hanya beberapa ulasan yang dapat ditemukan pasangan aspek dan opini yang sesuai berdasarkan data yang sudah di anotasi. Perhitungan akurasi berdasarkan presicion dan recall dari tiap-tiap data ulasan.

dapat dilihat sekilas tentang hasil presisi yang diperoleh dari rentang 0-1, dimana terdapat beberapa ulasan yang berhasil di ekstrak sempurna, namun jika terdapat banyak ulasan yang gagal ditemukan pasangan aspek dan opininya, dimana rata-rata dari *presicion* adalah 22% semantara untuk recal sebesar 21%. Berikut hasil lengkap perhitungan akurasi dengan data test dan data train.

|  |  |  |
| --- | --- | --- |
| Jenis Data | Presicion | Recall |
| Data Canon G3 | 22% | 21% |
| Data Test | 64% | 78% |

## Sentiment Classification

fiturnya, namun terdapat kalimat yang tidak dapat diketahui fiturnya dan terdapat kegagalan pada kalimat tertentu yang meng-ekstrak fitur lebih dari 1 sehingga dari ke 11 tersebut masih terdapat kegagalan, dengan nilai rata-rata recall sebesar 78% dan presisi sebesar 64%. Nilai yang terbilang cukup tinggi jika dibandingkan dengan hasil pengujian dengan data test. Hal ini diakibatkan pola kalimat dari datatest yang cenderung lebih sederhana dan tidak terlalu kompleks, kalimat yang sederhana cenderung dengan mudah.

Pengujian yang dilakukan dalam klasifikasi sentimen dengan menggunakan data Canon G3 dengan memanfaatkan label yang telah diberikan dan menghitung *presicion* dan *recall* dari hasil yang sudah didapatkan.