

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Penelitian Sebelumnya Yang Berkaitan**

Berikut ini adalah beberapa hasil penelitian yang terkait dan menjadi referensi pada penelitian ini :

1. Klasifikasi sentimen terdiri dari beberapa tahapan. Tahapan yang pertama adalah pemilihan model data. Tujuan dari pemilihan model data adalah agar terbentuk kombinasi terbaik antara data uji dan data latih. Pemilihan model data dilakukan menggunakan Stratified K-Fold Cross Validation, di mana dataset dibagi menjadi 2 fold agar proporsi dari data uji dan data latih adalah 50:50. Lalu tahapan selanjutnya adalah klasifikasi dengan memanfaatkan fungsi fit atau training. Proses fit tersebut memanfaatkan library MultinomialNB. Selanjutnya adalah melakukan prediksi terhadap data uji sehingga menghasilkan array yang berisi data uji beserta kelas prediksinya. Proses klasifikasi sentimen dilakukan sebanyak tiga kali pada setiap aspek. Terdapat dua dataset yang digunakan dalam proses klasifikasi sentimen yaitu dataset term weighting dan dataset yang berisi ulasan pelanggan beserta label kelas positif atau negatif.
2. Tabel perbandingan antara kelas yang sebenarnya (actual label) dengan hasil prediksi (prediction) metode Naïve Bayes dan SVM terhadap data testing. Dari hasil pengujian menunjukkan bahwa model yang menggunakan metode Naïve Bayes, memprediksi dengan benar lebih banyak sentimen dibandingkan metode SVM. Metode Naïve Bayes

memprediksi sebanyak 299 sentimen dari tweet dengan benar, lebih banyak 14 prediksi benar dibandingkan dengan metode SVM yang memprediksi 285. Dapat disimpulkan bahwa kedua metode dapat melakukan performa yang baik dan metode Naïve Bayes memiliki hasil akurasi, presisi, recall, dan F1-Score yang lebih baik dibandingkan dengan metode SVM.

3. Dalam penelitian ini, dilakukan pengujian menggunakan dua metode klasifikasi yaitu Naïve Bayes Classifier dan Support Vector Machine. Berdasarkan hasil pengujian, diperoleh beberapa hal yang menarik. Pertama, ketepatan klasifikasi menggunakan metode Naïve Bayes Classifier diperoleh dengan akurasi sebesar 81%, precision sebesar 80%, recall sebesar 99%, dan f1\_score sebesar 89%. Kedua, ketepatan klasifikasi menggunakan metode Support Vector Machine diperoleh dengan akurasi sebesar 87%, precision sebesar 88%, recall sebesar 96%, dan f1\_score sebesar 92%. Dari kedua metode tersebut, dapat disimpulkan bahwa Support Vector Machine lebih baik dalam mengklasifikasi data dengan nilai performansi yang lebih baik. Berdasarkan hasil analisis dan kesimpulan tersebut, para peneliti merekomendasikan agar pada penelitian selanjutnya dapat dilakukan pengambilan data dari media sosial populer lainnya seperti Facebook dan Instagram agar data yang didapat lebih bervariasi. Selain itu, dapat pula menggunakan metode klasifikasi yang lain agar diperoleh hasil klasifikasi yang lebih spesifik dan akurat.
4. Berdasarkan hasil penelitian analisis sentimen ini, dapat disimpulkan bahwa mayoritas netizen Indonesia memberikan tanggapan positif

terhadap kehadiran Non-Fungible Token di media sosial Twitter. Menurut penelitian ini, hasil positif ini disebabkan oleh banyaknya opini publik yang setuju dengan manfaat kehadiran Non-Fungible Token. Tentu saja, penelitian analisis sentimen ini masih memiliki kekurangan. Oleh karena itu, saran dari penulis untuk pengembangan selanjutnya adalah dengan mengumpulkan lebih banyak data tweet untuk klasifikasi, pada proses pelabelan pengguna lebih dapat memahami kosakata Indonesia yang baik dan benar, kemudian pada tahap pelabelan dapat dilakukan secara otomatis, pada pembersihan data perlu lebih diperhatikan, dan penelitian dapat dilakukan dengan metode lain untuk membandingkan dengan metode lain selain Support Vector Machine.

5. Dalam penelitian ini, hasil analisis sentimen masyarakat di Pulau Ambon terhadap Perusahaan Listrik Negara Cabang Ambon menunjukkan bahwa metode klasifikasi NBC cenderung menghasilkan sentimen positif sebesar 67%, sedangkan metode SVM lebih baik dengan tingkat akurasi sebesar 81.67%. Namun, kedua metode tersebut menghasilkan sentimen negatif yang tidak lebih dari 50%, sehingga dapat disimpulkan sentimen masyarakat di Pulau Ambon terhadap perusahaan listrik tersebut cukup baik. Hal ini juga dipengaruhi oleh layanan masyarakat, terutama akun Twitter @ambonlima yang memberikan informasi secara realtime. Penelitian juga menunjukkan bahwa proses pengujian pada K-Fold Cross Validation memiliki pengaruh dalam meningkatkan akurasi metode klasifikasi NBC dan SVM. Hasil penelitian tersebut menunjukkan bahwa algoritma yang berbeda pada proses klasifikasi dapat menghasilkan

prediksi yang berbeda, yang bisa benar atau salah pada kasus lain. Oleh karena itu, pengembangan selanjutnya dari penelitian ini perlu dilakukan untuk meningkatkan akurasi dari setiap metode klasifikasi yang digunakan, seperti dengan melakukan pengujian akurasi menggunakan n-gram dan memasukkan emoticon pada tweet sebagai bahan pertimbangan untuk representasi mood seseorang.

6. Dalam penelitian ini, penulis menguji empat model algoritma untuk mengklasifikasikan sentimen tweet terkait empat pasangan calon gubernur Jawa Barat periode 2018-2023. Algoritma yang diuji meliputi Naïve Bayes, Support Vector Machine, Naïve Bayes berbasis Genetic Algorithm, dan Support Vector Machine berbasis Genetic Algorithm. Data tweet yang digunakan dikelompokkan menjadi positif dan negatif dengan total data sebanyak 2643. Model yang diuji akan menghasilkan nilai akurasi, presisi, recall, dan AUC dari setiap algoritma. Hasil pengujian menunjukkan bahwa algoritma Support Vector Machine berbasis Genetic Algorithm adalah yang terbaik dengan rata-rata akurasi sebesar 93,03% dan AUC sebesar 0,869. Model ini dapat memberikan hasil terbaik dalam pengklasifikasian sentimen tweet terkait calon gubernur Jawa Barat periode 2018-2023.

## **2.2 Sekilas Tentang Review**

*Review* atau jika dalam bahasa Indonesia yaitu *reviu* menurut Kamus Besar Bahasa Indonesia (KBBI) adalah ulasan atau tinjauan terhadap sesuatu yang pernah dibaca, dilihat atau didengan tinjauan kembali atau ulangan.

### 2.3 Sekilas Tentang Analisa Sentimen

Opinion mining atau juga dikenal sebagai analisa sentimen adalah proses yang bertujuan untuk menentukan apakah polaritas kumpulan teks tulisan (dokumen, kalimat, paragraph, dll) cenderung ke arah positif, negatif, atau netral (Kontopoulos, Berberidis, Dergiades, & Bassiliades, 2013). Menurut (Medhat et al., 2014) menyatakan bahwa “analisa sentimen adalah teknik komputasi pendapat, perasaan dan subjektivitas teks”. Menurut Maynard dan Funk dalam (Medhat et al., 2014), teknik klasifikasi sentimen dapat terbagi atas:

1. *Machine Learning Approach* menerapkan algoritma machine learning terkenal dengan menggunakan fitur linguistik.
2. *Lexicon Based Approach* bergantung pada leksikon sentimen, sebuah kumpulan istilah sentimen yang dikenal dan dikompilasi. Hal ini dibagi menjadi dua, yaitu dictionary-based approach dan corpus-based approach yang menggunakan metode statistik atau semantik untuk mencari sentimen polaritas.
3. *Hybrid Approach*, adalah pendekatan hybrid menggabungkan kedua pendekatan leksikon sentimen sebagai peran kunci dalam sebagian besar metode

### 2.4 Sekilas Tentang Web Scraping

Web Scraping Turland (2010) adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya berupa halaman-halaman web dalam bahasa markup seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk

digunakan bagi kepentingan lain. Web scraping memiliki sejumlah langkah, sebagai berikut Josi, A. (2014):

1. *Create Scraping Template*: Pembuat program mempelajari dokumen HTML dari website yang akan diambil informasinya untuk tag HTML yang mengapit informasi yang akan diambil.
2. *Explore Site Navigation*: Pembuat program mempelajari teknik navigasi pada website yang akan diambil informasinya untuk ditirukan pada aplikasi web scraper yang akan dibuat.
3. *Automate Navigation and Extraction*: Berdasarkan informasi yang didapat pada langkah 1 dan 2 di atas, aplikasi web scraper dibuat untuk mengotomatisasi pengambilan informasi dari website yang ditentukan.
4. *Extracted Data and Package History*: Informasi yang didapat dari langkah 3 disimpan dalam tabel atau tabel-tabel database.

## **2.5 Sekilas Tentang Naïve Bayes Classifier**

Berikut penelitian terkait yang menjadi referensi pada penelitian ini

:

Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes.

### **2.5.1 Naïve Bayes Classifier Sebagai Metode Klasifikasi Analisa Sentimen**

Menurut Tseng et al. dalam (Dhande & Patnaik, 2014), Naive Bayes Classifier adalah model sederhana untuk klasifikasi. Model ini bekerja dengan baik untuk klasifikasi teks. Model ini merupakan bentuk sederhana dari Bayesian Network, dimana semua atribut independen diberi nilai kelas

variabel. Naive Bayes Classifier memiliki beberapa keunggulan seperti sederhana, cepat dan akurasi yang tinggi.

Naive Bayes merupakan klasifikasi paling sederhana dan paling umum digunakan. Naive Bayes menghitung probabilitas kelas berdasarkan distribusi kata-kata yang ada dalam dokumen (Medhat et al., 2014). Naive Bayes memiliki beberapa keunggulan seperti sederhana, cepat dan akurasi yang tinggi. Banyak peneliti telah melakukan klasifikasi sentimen dengan menggunakan Naive Bayes. Namun klasifikasi ini memiliki keterbatasan utama yang tidak mungkin selalu memenuhi asumsi independensi antara atribut. Dan ini mempengaruhi tingkat akurasi klasifikasi (Dhande & Patnaik, 2014).

Naive Bayes Classifier dapat dihitung menggunakan rumus :

$$P(c|term\ dokumen\ d) \\ = P(c) \times P(t_1|c) \times P(t_2|c) \times P(t_3|c) \times \dots \times P(t_n|c)$$

Keterangan :

$P(c)$  = Probabilitas prior dari kelas c

$t_n$  = Kata dokumen d ke-n

$P(c|term\ dokumen\ d)$  = Probabilitas suatu dokumen termasuk kelas c

$P(t_n|c)$  = Probabilitas suatu kata ke-n dengan diketahui kelas c

Probabilitas prior kelas c ditentukan dengan rumus :

$$P(c) = \frac{N_c}{N}$$

Keterangan :

$N_c$  = Jumlah kelas c pada seluruh dokumen

$N$  = Jumlah seluruh dokumen

Probabilitas kata ke-n ditentukan dengan menggunakan teknik laplacian smoothing:

$$P(tn|c) = \frac{count(tn, c) + 1}{count(c)|V|}$$

Keterangan :

$count(tn, c)$  = Jumlah term tn yang ditemukan di seluruh data latih dengan kategori c

$count(c)$  = Jumlah term di seluruh data latih dengan kategori c

$|V|$  = Jumlah seluruh term pada data latih

## 2.6 Sekilas Tentang Support Vector Machine (SVM)

Berikut penelitian terkait yang menjadi referensi pada penelitian ini :

Komparasi Algoritma Naive Bayes Dan Support Vector Machine Untuk Analisa Sentimen Review Film.

### 2.6.1 Support Vector Machine Sebagai Metode Klasifikasi Analisa

#### Sentimen

Support Vector Machines (SVM) telah menjadi metode klasifikasi dan regresi yang populer untuk masalah linear dan nonlinear. Keistimewaan dari Support Vector Machines berasal dari kemampuan untuk menerapkan pemisahan linear pada input data non linear berdimensi tinggi, dan ini diperoleh dengan menggunakan fungsi kernel yang diperlukan. Efektivitas Support Vector Machines sangat dipengaruhi oleh jenis fungsi kernel yang dipilih dan diterapkan berdasarkan karakteristik data (Haddi, Liu, & Shi, 2013). Banyak peneliti telah melaporkan bahwa Support Vector Machines metode yang paling akurat untuk teks klasifikasi (Moraes, Valiati, & Gavião Neto, 2013).

## **2.7 Pemrograman Aplikasi**

### **2.7.1 Website**

Menurut Rohi Abdullah (2018, h. 1) dalam bukunya yang berjudul “7 in 1 Pemrograman Web Untuk Pemula” web atau website dapat diartikan sebagai kumpulan halaman yang berisi informasi data digital baik berupa teks, gambar, animasi, suara dan video atau gabungan dari semuanya yang disediakan melalui jalur koneksi internet sehingga dapat diakses dan dilihat oleh semua orang di seluruh dunia.

Halaman website dibuat dengan menggunakan bahasa standar yaitu HTML. Skrip HTML ini akan diterjemahkan oleh web browser sehingga dapat ditampilkan dalam bentuk informasi yang dapat dibaca oleh semua orang. Rohi Abdullah juga membagi website dalam tiga jenis, yaitu website statis, website dinamis, dan website interaktif.

Dari penjelasan teori diatas, penulis menyimpulkan web adalah sebuah halaman informasi atau konten yang berkaitan dengan tujuan web tersebut dan dapat diakses oleh semua orang yang memiliki internet melalui web browser.

### **2.7.2 HTML**

Rohi Abdullah (2018, h. 7) mengklaim bahwa HTML merupakan singkatan dari Hypertext Markup Language yaitu bahasa standar web yang dikelola oleh W3C (World Wide Web Consortium) berupa tag-tag yang menyusun setiap halaman website yang menempatkan setiap elemen website sesuai layout yang diinginkan.

Tag adalah tanda awal dan tanda akhir yang digunakan sebagai pengapit suatu elemen. Elemen adalah nama penanda yang diapit oleh tag yang memiliki fungsi dan tujuan tertentu pada dokumen HTML. Atribut adalah properti elemen yang digunakan untuk mengkhususkan satu elemen.

### **2.7.3 PHP**

PHP (Hypertext Preprocessor) yaitu bahasa pemrograman web yang dapat disisipkan dalam skrip HTML dan bekerja di sisi server. Tujuan dari bahasa ini adalah membantu para pengembang web untuk membuat web dinamis dengan cepat. (Rohi Abdullah, 2018, h. 127)

Fungsi PHP dalam pemrograman web diantaranya sebagai berikut :

- a. Menghasilkan halaman dinamis pada website
- b. Membuat, membuka, menulis, membaca, menghapus, dan menutup file pada server
- c. Memproses data yang dikirim dari form
- d. Mengirim dan mengakses cookie
- e. Modifikasi data pada database

### **2.7.4 Cascading Style Sheet (CSS)**

CSS menurut Rohi Abdullah (2018, h. 45) adalah singkatan dari Cascading Style Sheet, yaitu dokumen web yang berfungsi mengatur elemen HTML dengan berbagai properti yang tersedia sehingga dapat tampil dengan berbagai gaya yang diinginkan.

CSS dapat mengendalikan ukuran gambar, warna, bagian tubuh pada teks, warna tabel, ukuran border, warna border, warna hyperlink, warna

mouse over, spasi antar paragraf, spasi antar teks, margin kiri, kanan, atas, bawah, dan parameter lainnya. CSS adalah bahasa style sheet yang digunakan untuk mengatur tampilan dokumen.

### **2.7.5 Javascript**

Rohi (2018, h. 193) mengatakan bahwa Javascript merupakan bahasa pemrograman web yang pemrosesannya dilakukan di sisi client. Javascript biasanya dijalankan ketika ada event tertentu yang terjadi pada halaman web baik event yang dilakukan oleh user, maupun event yang terjadi karena adanya perubahan pada halaman website.

Javascript berfungsi membuat sebuah halaman website lebih interaktif dan dinamis. Penggunaan kode javascript sangat dibutuhkan untuk website-website yang berorientasi pada kenyamanan pengakses (user-experience). Namun, penggunaan kode javascript ini tidak akan maksimal jika tidak dibarengi dengan penggunaan CSS. Jadi jika diibaratkan, HTML dan CSS adalah fisik tubuh dan kosmetik sedangkan javascript adalah nyawa dari tubuh tersebut.

### **2.7.6 Bootstrap**

Menurut Mohamad Rizki Adhiasta (2016) dalam jurnal “Perancangan Informasi Batik Solo Berbasis Bootstrap”. Bootstrap adalah sebuah framework yang dikembangkan pengembang Twitter pada pertengahan tahun 2010. Sebelum menjadi kerangka kerja open-source, Bootstrap dikenal sebagai Twitter Blueprint. Dan terus berkembang sampai saat ini dan Bootstrap telah menjadi salah satu front-end framework yang paling populer dan merupakan proyek open source di dunia. Platform ini

awalnya dikembangkan pada ajang Hackweek, sebuah perhelatan developer yang diadakan Twitter. Bootstrap digambarkan sebagai CSS sederhana namun dibangun dengan pre-processor yang menyediakan lebih banyak daya dan fleksibilitas ketimbang CSS standar. Awalnya dirilis pada Jumat, 19 Agustus, 2011, dan sudah memiliki lebih dari 20 produk release termasuk yang terbesar adalah versi 2 dan versi 3. Hadirnya Bootstrap 2, maka fungsionalitas responsif terhadap seluruh kerangka sebagai stylesheet opsional telah di tambahkan dan pada Bootstrap 3, maka responsif secara default telah ditambahkan untuk pendekatan pertamanya terhadap ponsel. Framework Bootstrap terdiri dari bahasa html dan css yang juga menyediakan efek javascript yang dibangun menggunakan jquery.

### **2.7.7 MySQL**

MySQL adalah Sistem Manajemen Database SQL Relasional Open Source yang paling populer. MySQL adalah salah satu RDBMS terbaik yang digunakan untuk mengembangkan berbagai aplikasi perangkat lunak berbasis web. MySQL dikembangkan, dipasarkan dan didukung oleh MySQL AB, yang merupakan perusahaan Swedia (Tutorials Point MySQL t.t.).

Secara garis besar fungsi dari MySQL adalah untuk membuat dan mengelola database pada sisi server yang memuat berbagai informasi dengan menggunakan bahasa SQL. Bahasa SQL ini memuat beberapa perintah seperti SELECT (mengambil), INSERT (menambah), UPDATE (mengubah), dan DELETE (menghapus).

MySQL juga memiliki beberapa kelebihan diantaranya :

- a. Tidak membutuhkan RAM besar
- b. MySQL dapat dipakai oleh beberapa user dalam waktu yang bersamaan (multiuser).
- c. MySQL bersifat open source. Lisensi yang dipakai adalah GPL dibawah pengelolaan Oracle.
- d. MySQL mempunyai fitur keamanan yang cukup baik. Ada beberapa lapisan keamanan yang diterapkan oleh MySQL, seperti level nama host, dan subnetmask.

## **2.8 Pemodelan Sistem Dengan UML**

Menurut Fowler mengklaim bahwa UML atau Unified Modelling Language adalah sebuah kumpulan notasi grafis yang membantu dalam mendeskripsikan dan merancang sistem perangkat lunak terutama sistem yang dibangun menggunakan gaya berorientasi objek (OO).

UML telah digunakan sebagai bahasa pemodelan tujuan umum (general-purpose) di bidang rekayasa perangkat lunak. Namun, sekarang UML telah banyak digunakan ke dokumentasi beberapa proses bisnis atau alur kerja. Misalnya, diagram aktivitas yang merupakan jenis diagram UML yang dapat digunakan sebagai pengganti diagram alur. Diagram aktivitas menyediakan cara pemodelan alur kerja yang lebih terstandarisasi serta berbagai fitur yang lebih luas untuk meningkatkan keterbacaan dan efektivitas.

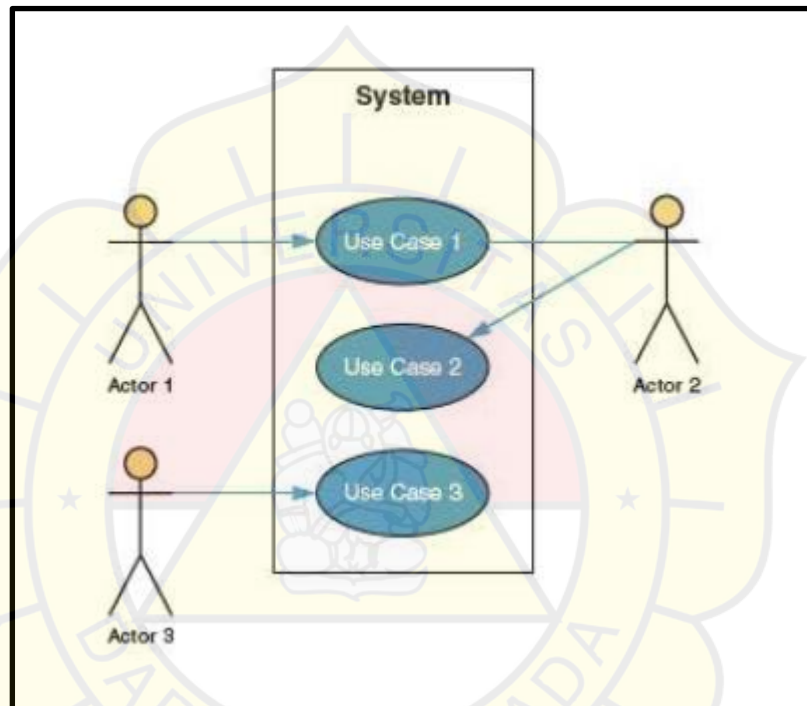
### **2.8.1 Use Case Diagram**

Menurut Fowler use case adalah sebuah teknik untuk menggambarkan kelayakan fungsional suatu sistem. Use case bekerja

dengan mendeskripsikan interaksi antara user dan sistem itu sendiri, serta memberikan narasi tentang bagaimana sistem itu digunakan.

Use Case menjelaskan bagaimana pengguna berinteraksi dengan sistem untuk melakukan beberapa aktivitas. Use Case juga digunakan untuk mengidentifikasi dan mengkomunikasikan persyaratan sistem.

Contoh Use Case Diagram :



**Gambar 2. 1** Gambar contoh use case



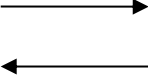




### 2.8.2 Activity Diagram

Activity diagram menurut Fowler (2004, h. 89) adalah teknik untuk menggambarkan logika prosedural, proses bisnis, dan alur kerja. Activity diagram mempunyai banyak kesamaan dengan flowchart, perbedaan utama keduanya adalah notasi dalam activity diagram mendukung perilaku paralel.

Activity diagram memiliki komponen diantaranya :

**Tabel 2. 1** Simbol activity diagram

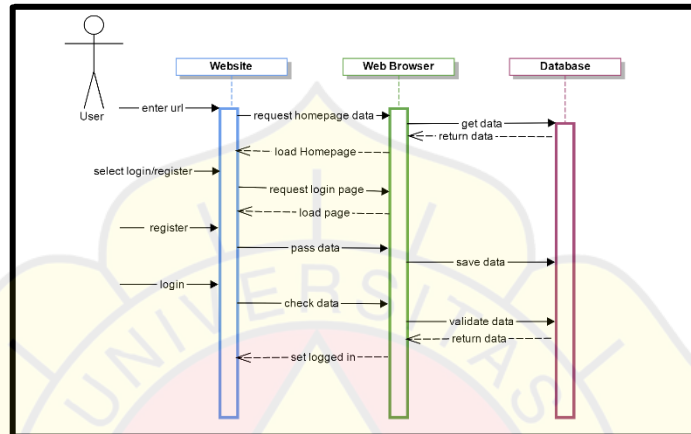
NO	SIMBOL	KETERANGAN

1		Menggambarkan awal sebuah proses
2		Menggambarkan sebuah kegiatan atau tugas yang perlu dilakukan
3		Menggambarkan sasaran yang mengawali kegiatan
4		Sebuah bar sinkronisasi. Simbol ini memperbolehkan anda untuk menggambarkan kegiatan yang dapat muncul secara paralel
5		Menggambarkan sebuah sasaran yang merupakan sebuah hasil dari kegiatan keputusan
6		Menggambarkan sebuah kegiatan keputusan / tindakan yang harus diambil pada kondisi tertentu
7		Menggambarkan akhir dari sebuah proses

### 2.8.3 Sequence Diagram

Sequence diagram menggambarkan interaksi antar objek di dalam dan di sekitar sistem (termasuk pengguna, display, dan sebagainya) berupa message yang digambarkan terhadap waktu. Sequence diagram terdiri atas dimensi vertikal (waktu) dan dimensi horizontal (objek-objek yang terkait). Sequence diagram biasa digunakan untuk menggambarkan skenario atau rangkaian langkah-langkah yang dilakukan sebagai respon dari sebuah event untuk menghasilkan output tertentu. Diawali dari apa yang men-trigger aktivitas tersebut, proses dan perubahan apa saja yang terjadi secara internal dan output apa yang dihasilkan. Masing-masing objek, termasuk

aktor, memiliki lifeline vertikal. Message digambarkan sebagai garis berpanah dari satu objek ke objek lainnya. Pada fase desain berikutnya, message akan dipetakan menjadi operasi/metoda dari class. Activation bar menunjukkan lamanya eksekusi sebuah proses, biasanya diawali dengan diterimanya sebuah message.



**Gambar 2. 2** Gambar contoh sequence diagram

#### 2.8.4 Deployment Diagram

Menurut (Irmayani & Susyatih, 2017) “Deployment Diagram menggambarkan hubungan antara software dan hardware terhadap sistem dan apa saja output yang dihasilkan. Software yang digunakan antara lainxampp sebagai server, phpmyadmin sebagai database dan NetBeans IDE 8.1 sebagai editor bahasa pemrograman java”.

Menurut Sukamto dan Shalahuddin (2015:154) memberikan batasan bahwa, Diagram deployment atau deployment diagram menunjukkan konfigurasi komponen dalam proses eksekusi aplikasi. Dapat disimpulkan bahwa Deployment Diagram menggambarkan hubungan antara software dan hardware terhadap sistem dan apa saja output yang dihasilkan dan

menunjukkan konfigurasi komponen dalam proses eksekusi aplikasi sebuah website atau aplikasi mobile.

