

## **BAB II**

### **LANDASAN TEORI**

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

Berikut adalah beberapa penelitian yang berkaitan dan menjadi referensi pada penulisan pada penelitian ini :

1. Dalam penelitian tersebut, dilakukan klasifikasi sentimen dengan menggunakan algoritma-algoritma tersebut, dan ditemukan bahwa hasil klasifikasi sentimen secara keseluruhan cenderung netral, dengan perbedaan yang tipis di setiap kategori. Akurasi tertinggi dicapai oleh SVM dengan tingkat keakuratan sebesar 76,5%, diikuti oleh NBC dengan tingkat keakuratan sebesar 72,3%, dan K-Nearest Neighbor (KNN) dengan tingkat keakuratan sebesar 59,1%. Meskipun demikian, penelitian ini juga mengungkapkan bahwa KNN memiliki potensi untuk memberikan analisis yang lebih mendalam dan akurat dengan menggunakan teknik yang berbeda. Dalam penelitian mendatang, dapat dilakukan eksplorasi lebih lanjut dan pengembangan teknik-teknik baru dalam menggunakan KNN untuk meningkatkan akurasi analisis sentimen. Selain itu, untuk meningkatkan hasil akurasi lebih lanjut dalam analisis sentimen, penelitian ini menyoroiti kebutuhan akan penggunaan metode lain seperti analisis emotikon dan alat-alat lainnya. Metode-metode ini dapat melengkapi algoritma-algoritma yang ada dan membantu dalam melakukan klasifikasi sentimen yang lebih baik.

2. Dalam percobaan dengan 3 kelas dan 16 kelas, ditemukan bahwa akurasi terbaik diperoleh dalam percobaan dengan 3 kelas. Pada percobaan ini, Naive Bayes Classifier (NBC) mencapai akurasi 86,11%, sedangkan K-Nearest Neighbor (KNN) mencapai akurasi 91,67%. Sedangkan pada percobaan dengan 16 kelas, NBC mencapai akurasi 25,20% dan KNN mencapai akurasi 22,05%. Dalam jumlah kelas yang sedikit, khususnya pada percobaan 3 kelas dengan jumlah data yang terbatas (120 data), KNN menghasilkan akurasi yang lebih tinggi daripada NBC. Namun, ketika jumlah kelas meningkat, seperti pada percobaan dengan 16 kelas dan jumlah data yang lebih banyak (422 data), NBC menghasilkan akurasi yang lebih baik daripada KNN. Oleh karena itu, dalam implementasi sistem menggunakan bahasa pemrograman Python, kami memilih algoritma KNN untuk percobaan dengan 3 kelas karena memiliki akurasi yang lebih tinggi. Pada algoritma KNN, ditemukan bahwa nilai  $k$  yang memberikan akurasi tertinggi adalah  $k=6$ ,  $k=7$ ,  $k=8$ , dan  $k=9$ . Dengan nilai  $k=7$ , diperoleh akurasi sebesar 91,67%, sedangkan untuk nilai  $k$  yang lain, akurasi tetap sebesar 86,11%. Hasil penelitian menunjukkan bahwa akurasi yang baik dapat dicapai jika sebaran data pada setiap kelas tidak memiliki rentang yang terlalu jauh, terutama pada percobaan dengan 3 kelas.
3. Berdasarkan penelitian yang dilakukan menggunakan bahasa pemrograman Python pada data tweet Twitter mengenai tanggapan masyarakat terhadap kebijakan PPKM, ditemukan hasil berikut. Jumlah data yang digunakan sebanyak 3.516. Dalam perbandingan akurasi

antara Naive Bayes Classifier (NBC) dan K-Nearest Neighbor (KNN) dengan menggunakan metode 10 K-Fold Cross Validation, ditemukan bahwa KNN memiliki akurasi terbaik sebesar 79,67%, sedangkan NBC memiliki akurasi sebesar 78,86%. Selanjutnya, dilakukan analisis terhadap polaritas sentimen masyarakat terhadap PPKM. Hasilnya menunjukkan bahwa sentimen positif mencakup 36,83% dari total data, dengan jumlah 1.295 tweet. Sentimen netral mencakup 54,15% dari total data, dengan jumlah 1.902 tweet. Sedangkan sentimen negatif mencakup 9,02% dari total data, dengan jumlah 317 tweet. Selain itu, ditemukan kata-kata yang muncul dengan frekuensi tertinggi dalam data tweet tersebut. Tiga kata dengan frekuensi tertinggi secara keseluruhan adalah "Covid", "Masyarakat", dan "Kota". Pada kelas sentimen positif, tiga kata dengan frekuensi tertinggi adalah "Ppkm", "Kota", dan "Covid". Pada kelas sentimen netral, tiga kata dengan frekuensi tertinggi adalah "Ppkm", "Level", dan "Panjang". Sedangkan pada kelas sentimen negatif, tiga kata dengan frekuensi tertinggi adalah "Ppkm", "Covid", dan "Panjang".

## **2.2 Sekilas Tentang Asuransi dan Asuransi Jiwa**

Menurut UU Nomor 2 Tahun 1992 tentang Usaha Perasuransian, menyebutkan bahwa usaha perasuransian yang sehat merupakan salah satu upaya untuk menanggulangi risiko yang dihadapi anggota masyarakat dan sekaligus merupakan salah satu lembaga penghimpun dana masyarakat, sehingga memiliki kedudukan strategis dalam pembangunan dan kehidupan perekonomian, dalam upaya memajukan kesejahteraan umum.

Sementara penjelasan lain nya mengenai Asuransi Jiwa menyebutkan, Perusahaan Asuransi Jiwa adalah perusahaan yang memberikan jasa dalam penanggulangan risiko yang dikaitkan dengan hidup atau meninggalnya seseorang yang dipertanggungkan.

## 2.3 Sekilas Tentang Naïve Bayes Classifier

### 2.3.1 Naïve Bayes Classifier Sebagai Metode Klasifikasi Text Mining

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).

Naïve Bayes Classifier dapat dihitung menggunakan rumus :

$$P(c|term\ dokumen\ d) \\ = P(c) \times P(t1|c) \times P(t2|c) \times P(t3|c) \times \dots \times P(tn|c)$$

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 :

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

Keterangan :

$N_c$  = Jumlah kelas c pada seluruh dokumen

$N$  = Jumlah seluruh dokumen

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

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

Keterangan :

$\text{count}(t_n, c)$  = Jumlah term  $t_n$  yang ditemukan di seluruh data latih dengan kategori c

$\text{count}(c)$  = Jumlah term di seluruh data latih dengan kategori c

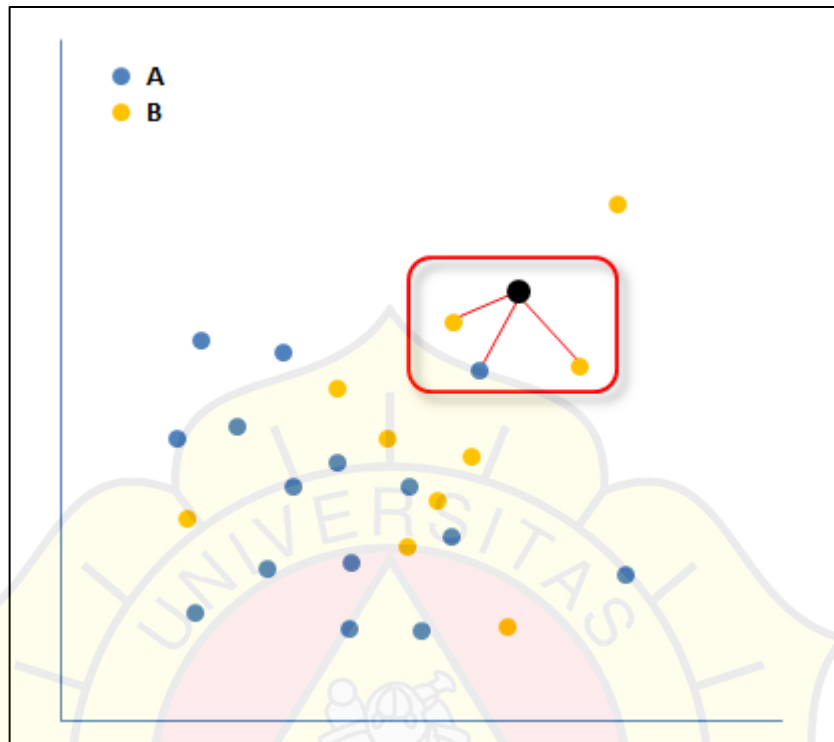
$|V|$  = Jumlah seluruh term pada data latih

## 2.4 Sekilas Tentang K-Nearest Neighbor (KNN)

### 2.4.1 K-Nearest Neighbor Sebagai Metode Klasifikasi Text Mining

K-Nearest Neighbor (KNN) adalah sebuah metode klasifikasi yang menggunakan K data terdekat sebagai acuan untuk menentukan kelas dari

data baru. Algoritma ini melakukan klasifikasi berdasarkan tingkat kemiripan atau kedekatan data tersebut dengan data lainnya.



**Gambar 2. 1** Gambar cara kerja metode KNN

Dari gambar 2.1, terdapat beberapa titik data yang terbagi menjadi dua kelas, yaitu A (biru) dan B (kuning). Jika kita memiliki data baru (hitam) yang perlu diprediksi kelasnya menggunakan algoritma KNN dengan nilai K sebesar 3, maka setelah menghitung jarak antara titik hitam tersebut dengan setiap titik data lainnya, kita dapat mengidentifikasi 3 titik terdekat. Dalam contoh ini, terdapat 2 titik kuning dan 1 titik biru yang terletak di dalam kotak merah. Oleh karena itu, kelas yang diprediksi untuk data baru (titik hitam) adalah B (kuning).

Dalam algoritma KNN, untuk menghitung jarak antara dua titik, kita menggunakan metode Euclidean Distance. Metode ini dapat digunakan baik dalam ruang 1 dimensi, 2 dimensi, maupun ruang multi-dimensi.

Dalam ruang 1 dimensi, perhitungan jarak dilakukan dengan menggunakan satu variabel bebas (independent variable). Dalam ruang 2 dimensi, terdapat dua variabel bebas yang digunakan. Sedangkan dalam ruang multi-dimensi, terdapat lebih dari dua variabel bebas yang terlibat dalam perhitungan jarak. Dengan menggunakan Euclidean Distance, kita dapat mengukur jarak antara dua titik dalam berbagai jenis ruang dimensi yang berbeda. Berikut adalah formula atau persamaan Euclidean distance 1-dimensional space :

$$dis = \sqrt{\sum(x_{1i} - x_{2i})^2}$$
$$dis = \sqrt{\sum(x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2 + \dots}$$

## 2.5 Pemrograman Aplikasi

### 2.5.1 Website

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.

### 2.5.2 HTML5

Pengembangannya dimulai sejak tahun 2004, ketika draf pertama disusun oleh Mozilla Foundation dan Opera Software (pengembang dua browser web populer). Namun, baru pada awal tahun 2013, draf akhirnya diserahkan kepada World Wide Web Consortium (W3C), badan pengatur internasional untuk standar web. Proses pengembangan HTML5 memakan

waktu beberapa tahun, tetapi saat ini kita sudah menggunakan versi yang sangat solid dan stabil, yaitu versi 5.1 (sejak 2016).

Namun, ini adalah siklus pengembangan yang tidak pernah berhenti, dan lebih banyak fungsionalitas pasti akan ditambahkan seiring waktu. Beberapa fitur terbaik dalam HTML5 untuk mengelola dan menampilkan media termasuk elemen `<audio>`, `<video>`, dan `<canvas>`, yang menambahkan suara, video, dan grafik canggih.

### 2.5.3 PHP

Bahasa pemrograman PHP menawarkan kemudahan dalam menyisipkan elemen dinamis ke dalam halaman web. Dengan memberikan ekstensi `.php` pada halaman, maka halaman tersebut akan langsung terhubung dengan bahasa pemrograman PHP. Ini memungkinkan pengembang untuk menulis kode dalam format seperti:

```
<?php
    echo " Today is " . date("l") . ". ";
?>
```

Baris kode di atas akan menginterpretasikan sebagai pesan yang menyatakan hari saat ini. Tanda pembuka `<?php` memberitahu server web untuk mengizinkan program PHP menginterpretasikan apa yang mengikuti hingga tanda penutup `?>`. Apa yang berada di luar konstruksi ini akan disampaikan secara langsung sebagai HTML ke klien. Misalnya, teks "Ini adalah berita terbaru." akan ditampilkan langsung di peramban.

Di dalam tag PHP, fungsi bawaan `date` digunakan untuk menampilkan hari berdasarkan waktu sistem server. Hasil dari kode di atas akan tampak seperti: “Hari ini adalah Rabu. Ini adalah berita terbaru.”

Ada beragam metode lain dalam memformat dan menampilkan informasi, yang akan dijelaskan lebih lanjut dalam bagian-bagian selanjutnya tentang PHP. PHP memberikan keuntungan bagi pengembang web, yaitu meskipun tidak secepat mengkompilasi kode dalam bahasa C atau bahasa serupa, PHP tetap cepat dan terintegrasi dengan markup HTML dengan sangat baik.

#### **2.5.4 PHP Machine Learning**

PHP Machine Learning (PHP-ML) adalah sebuah library yang dikembangkan untuk *handle* machine learning menggunakan PHP dan library ini termasuk algoritma machine learning dan API data *processing* yang dapat *handle* pembersihan data dan ekstraksi data. Machine learning yang berfokus pada penggunaan data dan algoritma untuk mengimitasi cara manusia belajar dan perlahan meningkatkan akurasi. PHP-ML ini membutuhkan PHP 7.1 atau lebih baru sebagai syarat agar dapat menjalankannya. PHP-ML dapat digunakan untuk beberapa metode seperti : Apriori, SVC, KNN, Naïve Bayes, Decision Tree.

Keuntungan menggunakan PHP Machine Learning :

1. Gratis dan mudah digunakan untuk aplikasi yang membutuhkan prediksi sederhana dan analisis data

2. Library ini memungkinkan para pengembang untuk melakukan tugas-tugas pembelajaran mesin yang kompleks dengan API yang sederhana dan intuitif.
3. Library ini membantu para developer web, yang merupakan pemula dalam bidang AI dan hanya ingin mempelajari konsep-konsep AI tanpa harus mempelajari Bahasa baru seperti R, python, atau lainnya.

### **2.5.5 Laravel**

Laravel adalah sebuah framework pengembangan website dengan syntax yang elegan dan ekspresif. Laravel bertujuan untuk mempermudah developer dengan mempermudah tugas-tugas umum seperti otentikasi, routing, session dan caching. Laravel juga bertujuan untuk membuat proses pengembangan menjadi menyenangkan tanpa mengorbankan fungsionalitas aplikasi.

### **2.5.6 CSS**

CSS merupakan pasangan yang penting bagi HTML, yang memastikan tata letak teks HTML dan gambar yang disisipkan secara konsisten dan sesuai dengan tampilan layar pengguna. Dengan adanya standar CSS3 dalam beberapa tahun terakhir, CSS kini menawarkan tingkat interaktifitas dinamis yang sebelumnya hanya didukung oleh JavaScript. Misalnya, tidak hanya dapat mengubah gaya elemen HTML untuk mengubah dimensi, warna, batas, jarak, dan lainnya, tetapi sekarang juga bisa menambahkan transisi dan transformasi animasi pada halaman web hanya dengan beberapa baris kode CSS. Penggunaan CSS bisa se-sederhana

memasukkan beberapa aturan antara tag `<style>` dan `</style>` di bagian head halaman web, seperti contoh berikut:

```
<style>
p {
    text-align: justify;
    font-family: Helvetica;
}
</style>
```

Aturan-aturan tersebut mengubah penjumlahan teks default dari tag `<p>` sehingga paragraf di dalamnya ditampilkan secara rata kanan-kiri dan menggunakan jenis huruf Helvetica. Seperti yang akan Anda pelajari di Bab 18, terdapat berbagai cara yang berbeda dalam menyusun aturan CSS, dan Anda juga dapat menyertakannya langsung di dalam tag atau menyimpan set aturan dalam file eksternal yang akan dimuat terpisah. Fleksibilitas ini tidak hanya memungkinkan Anda mengatur tampilan HTML dengan presisi, tetapi juga dapat memberikan fungsionalitas hover yang terintegrasi untuk menganimasikan objek saat kursor mouse melewatinya. Anda juga akan belajar bagaimana mengakses semua properti CSS elemen dari JavaScript maupun HTML.

### **2.5.7 Javascript**

JavaScript adalah teknologi dasar dalam pengembangan web yang memungkinkan interaksi pengguna yang dinamis dengan dokumen HTML. Dengan JavaScript, kita dapat melakukan berbagai tugas seperti validasi formulir dan menampilkan pesan interaktif. Namun, implementasi

JavaScript dapat bervariasi antara peramban web yang berbeda, sehingga dapat timbul masalah kompatibilitas.

Dalam melakukan penelitian ilmiah, penting untuk mempertimbangkan potensi batasan dan tantangan yang ditimbulkan oleh JavaScript. Masalah kompatibilitas dapat muncul akibat perbedaan implementasi JavaScript pada berbagai peramban, terutama pada peramban lama. Disinkronisasi ini dapat mempengaruhi keandalan dan reproduktibilitas eksperimen yang dilakukan di platform web.

Untuk memastikan keabsahan dan kekokohan penelitian ilmiah yang melibatkan JavaScript, peneliti perlu mempertimbangkan persyaratan kompatibilitas dari audiens target mereka. Mungkin diperlukan optimisasi kode atau eksplorasi pustaka dan teknik yang mengatasi masalah kompatibilitas. Dengan memperhatikan tantangan potensial ini dan mengadopsi strategi yang tepat, peneliti dapat mengurangi dampak kompatibilitas JavaScript terhadap keandalan dan generalisabilitas temuan ilmiah mereka.

## **2.6 Pemodelan Sistem Dengan UML**

### **2.6.1 Use Case Diagram**

Diagram use case adalah sebuah model yang digunakan untuk memvisualisasikan kebutuhan sistem secara keseluruhan pada tingkat yang lebih tinggi. Diagram use case digunakan untuk menggambarkan interaksi antara aktor dan use case yang terkait. Diagram ini membantu dalam memahami proses bisnis dan mempermudah komunikasi dengan para

pemangku kepentingan. Spesifikasi dan dokumentasi use case yang terdapat dalam diagram use case menjadi bagian inti dari pemodelan kebutuhan.

Diagram use case bersifat statis dan tidak menunjukkan aspek perilaku dari kebutuhan. Hubungan antara aktor dan use case tidak mencerminkan konsep waktu, sehingga diagram use case dikategorikan sebagai diagram statis. Namun, perlu diperhatikan bahwa diagram use case bukanlah representasi alur atau perilaku sistem. Alur proses akan dijelaskan secara detail dalam dokumentasi teks use case dan diagram aktivitas yang terkait. Dalam diagram use case, use case tidak dapat diuraikan seperti halnya diagram aliran data (DFD). Tidak terdapat tingkatan atau lapisan dalam diagram use case, karena semuanya berada pada level yang sama dalam model kebutuhan secara keseluruhan. Dokumentasi use case menjadi sumber yang penting dalam mengidentifikasi entitas bisnis yang kemudian akan menghasilkan kelas-kelas yang relevan.

Diagram use case mencakup aktor, use case, dan hubungan antara mereka. Aktor merupakan pihak eksternal yang berinteraksi dengan sistem, sedangkan use case adalah aksi atau fungsi yang dilakukan oleh sistem. Garis yang menghubungkan aktor dengan use case hanya menunjukkan asosiasi atau komunikasi, bukan ketergantungan atau aliran informasi. Untuk menjelaskan proses yang diinginkan dengan lebih jelas, seringkali diperlukan penjelasan tambahan, deskripsi, dan catatan dalam diagram use case.

## 2.6.2 Activity Diagram

Diagram aktivitas memodelkan alur atau proses dalam sebuah sistem. Oleh karena itu, diagram ini mirip dengan flowchart. Pemodelan alur ini dapat dilakukan pada level proses bisnis, dalam suatu use case, dan kadang-kadang antara use case yang berbeda. Aktivitas dalam diagram ini dapat berada pada level teknis yang detail atau pada level bisnis. Diagram aktivitas mendokumentasikan perilaku internal dalam use case, antara use case, atau keseluruhan proses bisnis.

Diagram aktivitas yang berada pada level yang lebih tinggi digunakan sebagai konteks untuk menunjukkan bagaimana berbagai proses bisnis saling terkait. Salah satu karakteristik penting dari diagram aktivitas adalah kemampuannya untuk menunjukkan ketergantungan antar aktivitas. Diagram aktivitas juga membantu dalam memetakan aktivitas-aktivitas tersebut ke aktor-aktor yang terkait dalam sistem. Selain itu, karena kemampuannya untuk menunjukkan alur sejalan (melalui fork dan join), diagram aktivitas juga dapat menunjukkan kejadian yang terjadi secara bersamaan dalam sistem. Kemampuan pemodelan multithreading yang disediakan oleh diagram aktivitas juga membantu dalam memodelkan ruang masalah.

Oleh karena itu, diagram ini menyediakan mekanisme yang sangat baik untuk memodelkan proses bisnis. Diagram aktivitas memiliki sifat perilaku. Hal ini dikarenakan diagram ini menunjukkan aktivitas-aktivitas dan urutan dimana aktivitas-aktivitas tersebut terjadi. Namun, diagram aktivitas tidak menunjukkan waktu secara spesifik kapan aktivitas-aktivitas

tersebut terjadi. Dalam hal ini, diagram aktivitas adalah flowchart perilaku yang bersifat umum. Oleh karena itu, diagram ini tidak termasuk dalam kategori diagram dinamis seperti diagram urutan. Sifat dari diagram aktivitas ini dapat dikategorikan sebagai diagram perilaku-statik.

### **2.6.3 Sequence Diagram**

Sequence diagram telah populer sejak Jacobson memperkenalkannya sebagai sarana untuk mendokumentasikan perilaku dalam use case. Dalam penggunaan sebelumnya, diagram urutan juga disebut sebagai diagram skenario, karena mereka menggambarkan, secara visual, suatu skenario (atau sebuah contoh) dalam sebuah use case. Karena kemampuannya yang praktis untuk menunjukkan apa yang terjadi "di dalam" sebuah use case, diagram urutan populer di kalangan analis bisnis dan desainer sistem. Setiap langkah dalam sebuah use case ditampilkan dalam diagram urutan sebagai catatan atau narasi. Sequence diagram mewakili interaksi terperinci antara aktor dan sistem, atau antara objek-objek yang bekerja sama dalam suatu rentang waktu tertentu. Informasi tentang apa yang terjadi sebelum interaksi dimulai dan setelah rentang waktu berakhir tidak ditampilkan secara langsung dalam diagram urutan. Meskipun demikian, diagram urutan memungkinkan kita untuk melihat urutan pesan yang dikirim antara objek-objek dan memahami bagaimana aliran informasi terjadi seiring waktu. Diagram urutan memberikan gambaran yang lebih dinamis tentang perilaku sistem dalam menjalankan use case.

#### 2.6.4 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*.