**USULAN PENELITIAN S2**

**DETEKSI BAHASA REKAM MEDIS DENGAN BIG DATA**

****

**Oleh:**

**LAKSITA KUSUMA WARDHANI**

**21/476211/PPA/06155**

**PROGRAM MAGISTER ILMU KOMPUTER**

**DEPARTEMEN ILMU KOMPUTER DAN ELEKTRONIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS GADJAH MADA**

**2018**

# HALAMAN PENGESAHAN USULAN PENELITIAN S2

Judul Tesis : Deteksi Bahasa Rekam Medis dengan Big Data

Nama Mahassiwa :

NIM :

Proposal telah diuji pada tanggal 29 Februari 2018 dan sudah diperbaiki sesuai saran penguji dan sudah disetujui para penguji.

|  |  |  |
| --- | --- | --- |
| **Nama Penguji** | | **Tanda Tangan** |
| 1. . | |  |
| 1. . | |  |
| 1. . | |  |
|  | Yogyakarta, ..........................2018 | |
| Mengetahui,  Pembimbing | Pengusul | |
| , Ph.D  NIP. | Nama Mahasiswa  NIM. | |

# DAFTAR ISI

Halaman

[HALAMAN PENGESAHAN USULAN PENELITIAN S2 ii](#_Toc95828987)

[DAFTAR ISI iii](#_Toc95828988)

[DAFTAR GAMBAR iv](#_Toc95828989)

[DAFTAR TABEL vi](#_Toc95828990)

[BAB I PENDAHULUAN 1](#_Toc95828991)

[1.1 Latar Belakang Masalah 1](#_Toc95828992)

[1.2 Rumusan Masalah 1](#_Toc95828993)

[1.3 Batasan Masalah 2](#_Toc95828994)

[1.4 Tujuan 2](#_Toc95828995)

[1.5 Manfaat 2](#_Toc95828996)

[BAB II TINJAUAN PUSTAKA 3](#_Toc95828997)

[BAB III LANDASAN TEORI 5](#_Toc95828998)

[3.1 Data Mining 5](#_Toc95828999)

[3.2 Data Collection 5](#_Toc95829000)

[3.3 Data Storing 5](#_Toc95829001)

[3.4 Data Cleaning 6](#_Toc95829002)

[3.5 Data Analysis 6](#_Toc95829003)

[3.6 Data Consumption 6](#_Toc95829004)

[BAB IV METODOLOGI PENELITIAN 7](#_Toc95829005)

[4.1 Studi Literatur 7](#_Toc95829006)

[4.2 Alat & Bahan 7](#_Toc95829007)

[4.3 Pengumpulan Data 7](#_Toc95829008)

[4.4 Prosedur Kerja 7](#_Toc95829009)

[BAB V JADWAL PENELITIAN 11](#_Toc95829010)

[DAFTAR PUSTAKA 12](#_Toc95829011)

# DAFTAR GAMBAR

[Gambar 3.1 Letak dan anatomi kelenjar tiroid 3](#_Toc514399651)

[Gambar 4.1 Rancangan sistem 5](#_Toc514399652)

[Gambar 4.6 Skema pengujian 6](#_Toc514399653)

# DAFTAR TABEL

[Tabel 2.1 Tinjauan pustaka 2](#_Toc514399658)

[Tabel 3.1 *Histologic classification of thyroid tumors (Reading et al., 2005)* 3](#_Toc514399659)

[Tabel 5.1 Jadwal penelitian 7](#_Toc514399660)

# BAB I PENDAHULUAN

## Latar Belakang Masalah

pengembangan teknologi yang semakin pesat membuat manusia mulai mengubah gaya hidup mereka dengan menggunakan teknologi, tidak dapat dipungkiri bahwa teknologi memiliki peran yang sangat besar bagi manusia. Teknologi tidak hanya mencakup pada aspek-aspek tertentu, dari kebutuhan pekerjaan hingga hiburan pun dipengaruhi oleh teknologi, seiring dengan perkembangan teknologi, maka semakin banyak pula data yang akan di dapatkan, data-data ini tidak akan berhenti berkembang seiring dengan penggunaan manusia terhadap teknologi, sebagai contoh data yang selalu bertambah dan berkembang setiap hari nya adalah rekam medis. Setiap hari nya rumah sakit memiliki data pasien yang baru datang maupun yang sedang diperiksa, data-data ini akan semakin bertambah dan semakin besar di untuk tangani, oleh karena nya dibutuhkan teknologi Big Data agar data besar yang di proses lebih terstruktur dan efektif. Dalam pengaplikasian Big Data ini digunakan sebuah data rekam medis yang akan dibaca deteksi bahasanya oleh pembacanya.

## Rumusan Masalah

Berdasarkan latar belakang yang dikemukakan di atas, maka rumusan masalah dalam penelitian ini sebagai berikut :

1. Apakah pemrosesan Big Data menghasilkan pemrosesan aplikasi lebih cepat
2. Apakah pengelolaan Big Data menghasilkan data yang bersih
3. Apakah penggunaan Big Data dapat membuat data lebih efektif

## Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini adalah sebagai berkut:

1. Fokus penelitian berfokus pada data informasi pasien yaitu rekam medis
2. Data informasi pasien atau rekam medis didapatkan dari www.pubmed.com

## Tujuan

Adapun tujuan penelitian ini adalah :

1. Untuk menjelaskan seberapa cepat aplikasi dalam pemrosesannya menggunakan Big Data
2. Untuk menjelaskan kebersihan data dalam pengelolaannya dengan Big Data
3. Untuk menjelaskan seberapa efektif penggunaan Big Data

## Manfaat

Manfaat dari penelitian ini berfokus pada penggunaan dan kemampuan Big Data dalam mengelola data rekam medis dengan jumlah yang banyak

# BAB II TINJAUAN PUSTAKA

Berberapa penelitian sebelumnya terkait dengan Big Data menurut (Tang et al., 2022) banyaknya teknik internet dan computer science menjadikan berbagai macam Big Data yang diperkenalkan dalam penelitian forecasting, dibawanya sebuah pengetahuan baru dan mengembangkan model prediksi. Pada penelitian ini dalam percobaan pertama di buat sebuah literature review dalam skala penelitian forecasting yang besar. Berdasarkan sumber Big Data, dalam penelitian forecasting ini menggunakan konten data berbasis user-generated (dari pengguna sosial media dalam bentuk teks, foto, dan lain-lain) data device-monitored (dari monitor meteorological, smart meters, GPS, dan lain-lain) dan aktivitas log data (untuk pencarian atau kunjungan web, online/offline marketing, pelayanan klinik, labolatorium, dan lain-lain). Perbedaan tipe data membawa informasi khusus dan karakteristik. Dominasi tugas forecasting yang berbeda, membutuhkan teknologi analisis yang berbeda dalam mengembangkan model forecasting. Sedangkan menurut (Park et al., 2015) data telah bertambah secara eksplosif karena mengembangan jaringan sosial dan *cloud computing.* Terdapat tantangan untuk menyimpan, memproses, menganalisis data yang besar, teknologi tradisional tidak akan menjadi solusi yang baik dalam memproses data yang besar. Sudah jelas bahwa *platform* big datamembantu pengguna dalam mengembangkan servis analisis secara efektif, bagaimanapun tetap dibutuhkan waktu yang lama untuk mengkoleksi data, mengembangkan algoritma, dan servis analisis. Di dalam penelitian ini dipersembahkan sebuah platform kolaborasi analisis big data untuk servis big data. *Developer* dapat berkolaborasi dengan satu sama lain pada platform untuk big data sebagai servis. *Developer* dapat berkolaborasi dengan satu sama lain pada platform melalui sharing data, algoritma, dan servis. Terdapat pula opini dari (Shen et al., 2022) bahwa personal big data dapat mempromosikan manjemen sosial dengan baik, aplikasi bisnis, dan servis personal, dan membawa kepastian benefit ekonomi dari pengguna. Kesulitan dari personal keamanan big data dan dalam keseimbangan diantara proteksi privasi data dan membagi premis dari kepuasan keamanan personal big data dan proteksi privasi, dalam paper ini, diusulkan metode harga personal big data berdasarkan *differential privacy* (PMDP) di desain dua mekanisme yang berbeda dari positif dan balik harga personal big data yang masuk akal. Di aplikasikan statistika aggregate pada dataset yang terbuka dan dievaluasi performanya dengan ekstensif, eksperimental menunjukan hasil bahwa PMDP dapat menyajikan harga personal big data yang masuk akal dan konpensasi yang adil pada pemilik data.

Dasar-dasar penelitian sebelumnya yang menjadi tinjauan pustaka pada penelitian ini dirangkum dalam Tabel 2.1.

Tabel 2.1 Tinjauan pustaka

| Nama | Penelitian | Metode | Hasil |
| --- | --- | --- | --- |
| Ling Tang, Jieyi Li, Hongchuan Du, Ling Li, Jun Wu, Shouyang Wang | Big Data in Forecasting Research: A Literature Review | Merievew tipe big data full scale pada penelitian forecasting | Beda tipe pada big data menghasilkan hotspot forecasting yang berbeda |
| Kyounghyun Park, Minh Chau Nguyen, Heesun Won | Web-based Collaborative Big Data Analytics on Big Data as a Service Platform | Mengkolaborasi platform big data analytic | Portal analytic terhubung dengan portal web servis dan support bermacam-macam manajemen big data dan alat development |
| Yuncheng Shena, Bing Guo, Yan Shen, Xuliang Duan, Xiangqian Dong, Hong Zhang, Chuanwu Zhang, Yuming Jiang | Personal big data pricing method based on differential privacy | Metode harga personal big data berdasarkan differential privacy (PMDP) | Semakin kecil perbedaan semakin baik perlindungan privasinya |

# BAB III LANDASAN TEORI

## Data Mining

Ada dua fokus persyaratan yaitu data extraction dan data mining, jika dijelaskan dengan sederhana data extraction adalah proses pengumpulan data dari halaman web ke database, selain itu terdapat data mining, data mining adalah proses dalam identifikasi pengetahuan yang sangat penting dari database, seperti data yang dikoleksi oleh data scientist. Seperti contoh pemilik toko e-commerce, setelah menggunakan berbagai macam teknik penelitian, dapat disimpulkan bahwa sekitar 70% orang menggunakan celana jeans. Ini disebut data extraction, lalu untuk mempelajari lebih dalam, dapat dipelajari data umur dan jenis kelamin yang menggunakan Brand 1 dan Brand 2. Proses ini dinamakan data mining. Berberapa alat data mining yang berguna yaitu RapidMiner, TeraData, dan Kaggle.

## Data Collection

Big data tidak memiliki tombol “berhenti”, selama dunia berkembang, data akan tetap mengikuti perkembangannya. Data butuh di ekstraksi secara terus menerus, seperti contoh, akan ada orang yang menggunakan Brand 1 dan menggantinya ke Brand 2 dan seterusnya, kemungkinannya sangat tak terhingga. Ekstraksi data akan lebih mudah menggunakan alat seperti import.io

## Data Storing

Jika membayangkan Google maka akan banyak data yang disimpan, tentunya tidak menggunakan sistem yang tradisional seperti file, CD, DVD, dan lain-lain. Tipe penyimpanan yang akan digunakan bergantung pada skala dari sebuah bisnis, sistem kapasitas data baik menggunakan infrastruktur yang menggunakan alat analitik terbaru dan tempat penyimpanan data. Data dapat disimpan pada tempat penyimpanan data seperti Cloudera, Hadoop, and Talend, tempat penyimpanan data adalah satu langkah yang dapat diaplikasikan diantara berberapa langkah lainnya.

## Data Cleaning

Data dapat datang dalam berbagai macam bentuk dan tingkatan, berberapa ada yang baik dan tidak baik terutama yang di ekstrasi dari web. Sebelumnya seluruh data yang di esktraksi butuh untuk dibersihkan, dalam proses pembersihan, semua data yang tidak dibutuhkan maupun tidak akurat akan di filter, setelah proses ini, data yang ada adalah data yang ingin di fokuskan, pembersihan mendukung untuk menstrukturkan data dengan baik, sebagai contoh, diketahui angka dan tipe orang yang menggunakan celana jeans, selagi di bersihkan, maka dapat dihapuskan duplikasi data, data yang salah, area atau informasi yang tidak diinginkan dan lain sebagainya, dapat menggunakan Data Cleaner atau OpenRefine untuk tujuan ini.

## Data Analysis

Peran terbesar dari big data adalah analisis, ketika menganalisis data datang dari pola pelanggan, tingkah laku, dan lain-lain. Metode penelitian penyelidikan menunjukan big data analisis yang sangat membantu, analisis adalah tentang menanyakan pertanyaan yang spesifik dan mencari jawaban untuk itu. Qubole dan Statwing adalah alat analisis data yang sangat membantu seperti contoh seperti pertanyaan apakah pembeli menyukai jeans berkantung dua? Warna apa yang menjadi kesukaan mereka dan lain-lain.

## Data Consumption

Data yang dikonsumsi meliputi : Klasifikasi trend penjualan dalam pasar, yang mana pasar tersebut dapat mempromosikan produk yang paling laku terjual. Digunakan oleh badan pemerintahan yang mana digunakan untuk meraih demografi, geografi, dan etnik yang tepat. Penjual menemukan big data sangatlah berguna untuk mencari iklan mana yang tepat untuk produk mereka.

# BAB IV METODOLOGI PENELITIAN

## Studi Literatur

Studi literatur dilakukan dengan mempelajari penelitian sebelumnya yang terkait yang diperoleh dari berbagai macam sumber seperti buku, artikel, publikasi jurnal, paper, tesis, disertasi dll.

## Alat & Bahan

#### Alat

Alat yang digunakan dalam penelitian ini yaitu:

1. PC/Laptop dengan spesifikasi *processor* Ryzen 7 CPU @2,5 GHz, RAM 8 GB, sistem operasi *Windows* 11 64 bit.
2. Mapreduce Hadoop 1.2.1
3. Apache Kafka

#### Bahan

Bahan yang digunakan dalam penelitian ini adalah data medis.

## Pengumpulan Data

Data diambil dari rekam medis www.pubmed.com

## Prosedur Kerja

##### Analisis dan perancangan sistem

Analisis big data adalah proses mencari pola, trend, dan hubungan pada dataset masif yang tidak dapat dilakukan oleh tenik dan alat manajemen data tradisional. Cara yang terbaik dalam memahami analisis big data adalah membandingkannya dengan data analitik biasa.

Pada pendekatan tradisional, biasanya analisis mengambil tempat pada waktu tertentu atau acara, dibandingkan dengan big data, analisis biasanya terjadi secara *real time* dan dapat penemuan dapat diketahui secara cepat.

Data dihasilkan dari berbagai macam sumber termasuk sensor, *log files,* dan media sosial dapat dimanfaatkan dengan baik dengan mandiri ataupun pelengkap data transaksional yang sudah dimiliki oleh organisasi

##### Business Problem Definition

Mempelajari domain bisnis untuk menyajikan motivasi dan tujuan untuk melaksanakan analisis, pada tahap ini masalah sudah diidentifikasi dan asumsi akan membuat seberapa besar potensi bertambah pada rumah sakit setelah melaksanakan analisis, hal penting lainnya adalah membuat masalah bisnis sebagai tantangan analisis yang dapat dialamatkan menjadi fase subsequent. Selain itu dapat ditentukan apakah masalah yang diidentifikasi merupakan masalah big data atau bukan berdasarkan kebutuhan bisnis dalam kasus bisnis, dan menjadi syarat dalam big data memilih karakteristik volume, kecepatan, dan variasi

##### Data Definition

Setelah kasus bisnis telah diidentifikasi, sekarang saatnya untuk mencari kumpulan data yang sesuai untuk dikerjakan, analisis dilakukan untuk melihat apa yang dilakukan rumah sakit lain pada kasus serupa. Bergantung pada kasus bisnis dan lingkup analisis yang ditangani. Sumber dataset dapat berupa internal dan eksternal rumah sakit, kumpulan data internal di dapat dari sumber internal seperti form *feedback*, form perangkat lunak yang ada, sedangkan data eksternal di dapat dari kumpulan data pihak ketiga.

##### Data Acquisition dan Filtration

Setelah sumber data teridentifikasi, kini saatnya mengumpulkan data dari sumber tersebut. Jenis data ini sebagian besar tidak terstruktur. Kemudian menjadi sasaran penyaringan, seperti penghapusan data yang rusak atau data yang tidak relevan, yang tidak memiliki cakupan untuk tujuan analisis. Di sini data yang rusak berarti data yang mungkin memiliki catatan yang hilang, atau yang menyertakan tipe data yang tidak kompatibel. Setelah penyaringan, salinan data yang disaring disimpan dan dikompresi, karena dapat digunakan di masa mendatang, untuk beberapa analisis lainnya.

##### Data Extraction

Sekarang data difilter, tetapi mungkin ada kemungkinan bahwa beberapa entri data mungkin tidak kompatibel, untuk memperbaiki masalah ini, fase terpisah dibuat, yang dikenal sebagai fase ekstraksi data. Pada fase ini, data yang tidak sesuai dengan ruang lingkup analisis yang mendasarinya, diekstraksi dan ditransformasikan dalam bentuk seperti itu.

##### Data Munging

Data dikumpulkan dari berbagai sumber, sehingga data menjadi tidak terstruktur. Mungkin ada kemungkinan, bahwa data tersebut mungkin memiliki kendala, yang tidak sesuai, yang dapat menyebabkan hasil yang salah. Oleh karena itu ada kebutuhan untuk membersihkan dan memvalidasi data. Ini termasuk menghapus data yang tidak valid dan menetapkan aturan validasi yang kompleks. Ada banyak cara untuk memvalidasi dan membersihkan data. Misalnya, kumpulan data mungkin berisi beberapa baris, dengan entri nol. Jika ada kumpulan data yang serupa, maka entri tersebut akan disalin dari kumpulan data itu, jika tidak, baris tersebut akan dihapus.

##### Data Aggregation dan Representation

Data dibersihkan dan divalidasi, terhadap aturan tertentu yang ditetapkan oleh rumah sakit. Tetapi data mungkin tersebar di beberapa kumpulan data, dan tidak disarankan untuk bekerja dengan banyak kumpulan data. Oleh karena itu, kumpulan data digabungkan. Contoh: Jika ada dua set data, yaitu bagian Akademik Siswa dan bagian Data Pribadi Siswa, maka keduanya dapat digabungkan melalui bidang yang sama, yaitu nomor gulungan. Fase ini membutuhkan operasi intensif karena jumlah data bisa sangat besar. Otomasi dapat menjadi pertimbangan, sehingga hal-hal ini dijalankan, tanpa campur tangan manusia.

##### Exploratory Data Analysis

Di sinilah langkah sebenarnya, tugas analisis. Tergantung pada sifat dari masalah data besar, analisis dilakukan. Analisis data dapat diklasifikasikan sebagai analisis Konfirmasi dan analisis Eksplorasi. Dalam analisis konfirmatori, penyebab suatu fenomena dianalisis terlebih dahulu. Asumsi itu disebut hipotesis. Data dianalisis untuk menyetujui atau menolak hipotesis. Analisis semacam ini memberikan jawaban pasti atas beberapa pertanyaan spesifik dan menegaskan apakah suatu asumsi itu benar atau tidak. Dalam analisis eksplorasi, data dieksplorasi untuk memperoleh informasi, mengapa suatu fenomena terjadi. Jenis analisis ini menjawab “mengapa” suatu fenomena terjadi. Analisis semacam ini tidak memberikan definitif, sementara itu memberikan penemuan pola.

##### Data Visualization

Sekarang kita memiliki jawaban untuk beberapa pertanyaan, menggunakan informasi dari data dalam kumpulan data. Tetapi jawaban ini masih dalam bentuk yang tidak dapat disajikan kepada pengguna bisnis. Semacam representasi diperlukan untuk memperoleh nilai atau beberapa kesimpulan dari analisis. Oleh karena itu, berbagai alat digunakan untuk memvisualisasikan data dalam bentuk grafik, yang dapat dengan mudah ditafsirkan oleh pengguna bisnis. Visualisasi dikatakan mempengaruhi interpretasi hasil. Selain itu, memungkinkan pengguna untuk menemukan jawaban atas pertanyaan yang belum dirumuskan.

##### Utilization of analysis results

Analisis selesai, hasilnya divisualisasikan, sekarang saatnya pengguna bisnis mengambil keputusan untuk memanfaatkan hasilnya. Hasilnya dapat digunakan untuk optimasi, untuk menyempurnakan proses bisnis. Hal ini juga dapat digunakan sebagai masukan untuk sistem untuk meningkatkan kinerja.

# DAFTAR PUSTAKA

Park, K., Nguyen, M. C., & Won, H. (2015). Web-based collaborative big data analytics on big data as a service platform. *International Conference on Advanced Communication Technology, ICACT*, *2015*-*August*, 564–567. https://doi.org/10.1109/ICACT.2015.7224859

Shen, Y., Guo, B., Shen, Y., Duan, X., Dong, X., Zhang, H., Zhang, C., & Jiang, Y. (2022). Personal big data pricing method based on differential privacy. *Computers and Security*, *113*, 102529. https://doi.org/10.1016/j.cose.2021.102529

Tang, L., Li, J., Du, H., Li, L., Wu, J., & Wang, S. (2022). Big Data in Forecasting Research: A Literature Review. *Big Data Research*, *27*, 100289. https://doi.org/10.1016/j.bdr.2021.100289