Big Data Analytics untuk Meningkatkan Aksesibilitas: Studi Kasus tentang Pemrosesan Kumpulan Data Penyandang Disabilitas Menggunakan Metode Linear Agression

*Big Data Analytics for Enhancing Accessibility: A Case Study on Processing Datasets of People with Disabilities*

**1Habibie Ed Dien\*, 2 Muhammad As’ad Prabowo, 3Nafarul Hamka**

1,2,3Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang

Jl. Soekarno Hatta No. 9 Kota Malang, Indonesia

\*e-mail: [*habibie@polinema.ac.id*](mailto:habibie@polinema.ac.id)[*mailto:rozalitoyib@umb.ac.aid*](mailto:rozalitoyib@umb.ac.aid)

**Abstrak**

Penelitian ini membahas penerapan Big Data Analytics untuk meningkatkan aksesibilitas bagi penyandang disabilitas melalui studi kasus pemrosesan kumpulan data penyandang disabilitas. Aksesibilitas yang lebih baik bagi penyandang disabilitas adalah hal yang penting untuk memastikan bahwa mereka dapat mengakses layanan dan sumber daya yang diperlukan untuk hidup mandiri. Dalam penelitian ini, kami mengidentifikasi dan menganalisis berbagai sumber data yang relevan, termasuk data demografis, data kesehatan, data transportasi, dan data infrastruktur. Kami menggunakan metode analisis Big Data Analytics untuk menggali wawasan berharga dari kumpulan data tersebut.

Melalui tinjauan literatur, kami menemukan bahwa aksesibilitas bagi penyandang disabilitas masih menghadapi tantangan yang signifikan. Namun, penggunaan Big Data Analytics telah terbukti menjadi alat yang berpotensi dalam mengatasi masalah ini. Dengan memanfaatkan potensi Big Data Analytics, kita dapat mengidentifikasi pola, tren, dan wawasan yang berkaitan dengan aksesibilitas, dan mengembangkan solusi yang lebih efektif.

Hasil penelitian ini memberikan pemahaman yang lebih baik tentang tantangan dan peluang dalam meningkatkan aksesibilitas bagi penyandang disabilitas. Diharapkan penelitian ini dapat memberikan rekomendasi kebijakan dan praktik terbaik untuk menciptakan masyarakat yang lebih inklusif, di mana semua individu, tanpa memandang keterbatasan yang mereka miliki, memiliki akses yang setara terhadap layanan dan sumber daya yang diperlukan.

**Kata kunci:** Big Data Analytics, aksesibilitas, penyandang disabilitas, pemrosesan data, studi kasus.

***Abstract***

*This research discusses the application of Big Data Analytics to improve accessibility for persons with disabilities through case studies of processing data sets of persons with disabilities. Better accessibility for persons with disabilities is important to ensure that they can access the services and resources needed to live independently. In this research, we identify and analyze a variety of relevant data sources, including demographic data, health data, transportation data, and infrastructure data. We use the Big Data Analytics analysis method to extract valuable insights from these data sets.*

*Through a review of the literature, we found that accessibility for persons with disabilities still faces significant challenges. However, the use of Big Data Analytics has proven to be a potential tool in addressing this problem. By tapping into the potential of Big Data Analytics, we can identify patterns, trends, and insights related to accessibility, and develop more effective solutions.*

*The results of this study provide a better understanding of the challenges and opportunities in improving accessibility for persons with disabilities. It is hoped that this research can provide policy recommendations and best practices for creating a more inclusive society, where all individuals, regardless of their limitations, have equal access to the necessary services and resources.*

***Keywords:*** *Big Data Analytics, accessibility, persons with disabilities, data processing, case studies.*

# Pendahuluan (or Introduction)

Studi ini membahas penerapan Big Data Analytics untuk meningkatkan aksesibilitas bagi penyandang disabilitas. Aksesibilitas yang lebih baik bagi penyandang disabilitas adalah hal yang sangat penting untuk memastikan bahwa mereka dapat mengakses layanan dan sumber daya yang diperlukan untuk hidup mandiri. Namun, masih ada tantangan yang signifikan dalam mencapai aksesibilitas yang memadai. Dalam konteks ini, Big Data Analytics menawarkan potensi besar dalam mengidentifikasi pola dan tren yang berkaitan dengan aksesibilitas serta memberikan wawasan yang berharga untuk meningkatkan kondisi tersebut.

Penelitian ini akan menggunakan studi kasus tentang pemrosesan kumpulan data penyandang disabilitas yang diperoleh dari berbagai sumber. Data tersebut meliputi data demografis, data kesehatan, data transportasi, dan data infrastruktur. Dengan memanfaatkan analisis Big Data, kita dapat menggali informasi yang signifikan dan mengidentifikasi area yang membutuhkan perhatian lebih dalam meningkatkan aksesibilitas bagi penyandang disabilitas.

Melalui penelitian ini, diharapkan akan ditemukan solusi efektif dan rekomendasi kebijakan untuk meningkatkan aksesibilitas bagi penyandang disabilitas. Dengan memanfaatkan potensi Big Data Analytics, kita dapat mencapai masyarakat yang lebih inklusif dan memberikan peluang yang setara bagi penyandang disabilitas, sehingga mereka dapat berpartisipasi secara penuh dalam kehidupan sosial, ekonomi, dan kesehatan.

# Tinjauan Literatur (or Literature Review)

**Spark**

Spark atau dikenal juga dengan nama Apache Spark merupakan open-source cluster framework computing, yang dibangun untuk pemrosesan big data dengan cepat, yang pada awalnya dikembangkan oleh AMPLab di UC Berkeley. Apache spark memiliki algoritma yang berbeda dengan Map/Reduce, tetapi dapat berjalan diatas Hadoop melalui YARN. Spark menyediakan API pemrograman Java, Python, Scala, dan SQL, dan dapat digunakan untuk menjalankan berbagai jenis proses secara efisien, termasuk proses ETL, data streaming, machine learning, perhitungan grafik, dan SQL.

**Google Colab**

Google Colaboratory atau Google Colab adalah executable document yang memungkinkan kamu dalam menulis, mengedit, serta membagikan program yang sudah disimpan pada drive maupun yang baru kamu buat. Selama menggunakan tools ini, kamu tidak perlu menginstal software tambahan lagi. Kamu hanya perlu akun Google untuk login dan menyimpan file Python ke Google Drive milikmu.

Tools ini sangat cocok digunakan bagi programmer atau pemula yang ingin belajar python. Selain itu, seperti namanya, Google Colab juga mendukung fitur kolaborasi dengan tim. Nantinya notebook atau dokumen yang digunakan untuk menulis program dapat diedit berbarengan oleh tim kamu seperti halnya mengedit pada Google Docs.

Jika sebelumnya kamu pernah mendengar Jupyter Notebook yang dijalankan pada browser, pada dasarnya Google Colab mirip dengan itu. Tools ini memungkinkan kamu untuk menjalankan program python tanpa perlu repot melakukan instalasi atau mengunduh text editor terlebih dulu. Justru, seluruh setup sepenuhnya diserahkan pada cloud.

**Linear Regression**

Linear Regression (Regresi Linear) adalah suatu regresi linear yang digunakan untuk mengestimasi atau memprediksi hubungan antara dua variabel dalam penelitian kuantitatif. Dimana regresi linear ini mampu membuat satu asumsi tambahan yang mengkorelasikan antara variabel independen dan dependen melalui garis yang paling sesuai dari titik data garis lurus. Artinya, bukan kurva atau semacam faktor pengelompokkan.

Meski demikian, regresi linear memiliki keterbatasan, karena dalam data terbaik pun tidak menceritakan kisah yang lengkap. Analisis regresi biasanya digunakan dalam penelitian untuk menetapkan bahwa ada korelasi antar variabel.

Namun, korelasi yang tidak sama sebab akibat adanya hubungan antara dua variabel tidak berarti yang satu menyebabkan yang lainnya terjadi. Bahkan garis dalam regresi linear sederhana yang cocok dengan titik data mungkin tidak menjamin hubungan sebab-akibat.

Menurut Sugiyono (2011), regresi linear sederhana adalah regresi linear yang didasarkan pada hubungan fungsional maupun kausal satu variabel independen dengan satu variabel dependen.

# Metode Penelitian (or Research Method)

**Big data lifecycle Analytics**



Big data lifecycle merupakan serangkaian tahapan yang menggambarkan alur data besar (big data) dari pengumpulan awal hingga pemanfaatan dan visualisasi akhir. Berikut adalah penjelasan mengenai tahapan dalam big data lifecycle hingga tahap visualisasi:

* **Pengumpulan Data:** Tahap pertama dalam big data lifecycle adalah pengumpulan data. Data dapat berasal dari berbagai sumber, seperti sensor, perangkat lunak, basis data, media sosial, dan lain-lain. Pada tahap ini, data dikumpulkan dan disimpan dalam format yang sesuai.
* **Penyimpanan Data:** Setelah dikumpulkan, data disimpan dalam sistem penyimpanan yang sesuai, seperti basis data, data warehousing, atau sistem file terdistribusi. Penyimpanan data yang efisien dan dapat diakses dengan cepat menjadi penting karena volume data yang besar.
* **Pengolahan Data:** Setelah data disimpan, tahap berikutnya adalah pengolahan data. Ini melibatkan transformasi, pembersihan, dan pemodelan data. Data yang tidak terstruktur atau tidak teratur akan diperbaiki dan dibentuk menjadi format yang lebih terstruktur dan dapat dimanfaatkan.
* **Analisis Data:** Setelah data diolah, tahap berikutnya adalah analisis data. Analisis data melibatkan penggunaan teknik dan algoritma analitis untuk mendapatkan wawasan, pola, dan tren dari data. Analisis ini dapat melibatkan teknik seperti analisis statistik, machine learning, atau data mining.
* **Penyajian Data:** Setelah analisis data, hasilnya perlu disajikan dengan cara yang dapat dipahami oleh pengguna. Tahap penyajian data melibatkan transformasi hasil analisis menjadi format yang lebih terstruktur, misalnya dalam bentuk laporan, tabel, atau grafik. Tujuan utamanya adalah untuk menyampaikan informasi yang relevan dan mudah dipahami oleh pemangku kepentingan.
* **Visualisasi Data:** Tahap terakhir dalam big data lifecycle adalah visualisasi data. Visualisasi data melibatkan representasi grafis data menggunakan berbagai jenis grafik, diagram, peta, dan visualisasi interaktif lainnya. Tujuan utamanya adalah untuk memperjelas dan memvisualisasikan pola, tren, dan informasi yang terkandung dalam data. Visualisasi data dapat membantu pemangku kepentingan untuk dengan mudah menginterpretasikan dan mengambil keputusan berdasarkan data yang kompleks.

Dalam keseluruhan big data lifecycle, tahap visualisasi merupakan salah satu tahap yang penting, karena dapat membantu mengkomunikasikan wawasan dan informasi yang diperoleh dari analisis data kepada pemangku kepentingan. Dengan menggunakan visualisasi yang tepat, pemangku kepentingan dapat dengan cepat memahami data yang kompleks dan mengambil keputusan yang lebih baik berdasarkan informasi yang diberikan.

**3.1 Pengumpulan data**

Kami mengambil dataset mengenai data penyandang disabilitas di provinsi jawabarat dari website opendatajabar.

Dataset ini berisi data jumlah penduduk penyandang disabilitas berdasarkan kategori disabilitas di Provinsi Jawa Barat dari tahun 2013 s.d. 2021.

Dataset terkait topik Kependudukan ini dihasilkan oleh Dinas Kependudukan dan Pencatatan Sipil yang dikeluarkan dalam periode 1 tahun sekali.

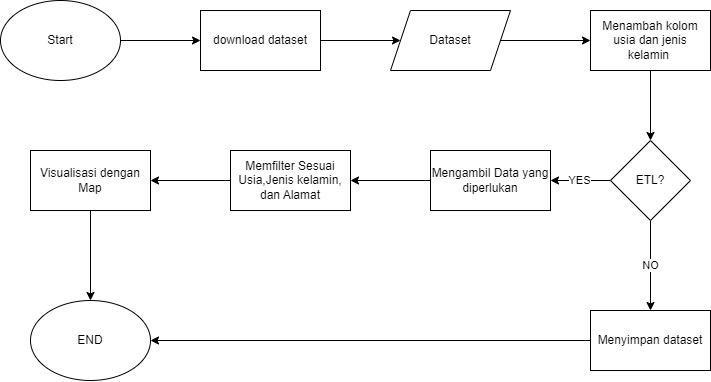
Penjelasan mengenai variabel di dalam dataset ini:

* kode\_provinsi: menyatakan kode Provinsi Jawa Barat sesuai ketentuan BPS merujuk pada aturan Peraturan Badan Pusat Statistik Nomor 3 Tahun 2019 dengan tipe data numerik.
* nama\_provinsi: menyatakan lingkup data berasal dari wilayah Provinsi Jawa Barat sesuai ketentuan BPS merujuk pada aturan Peraturan Badan Pusat Statistik Nomor 3 Tahun 2019 dengan tipe data teks.
* kode\_kabupaten\_kota: menyatakan kode dari setiap kabupaten dan kota di Provinsi Jawa Barat sesuai ketentuan BPS merujuk pada aturan Peraturan Badan Pusat Statistik Nomor 3 Tahun 2019 dengan tipe data numerik.
* nama\_kabupaten\_kota: menyatakan lingkup data berasal dari setiap kabupaten dan kota di Provinsi Jawa Barat sesuai penamaan BPS merujuk pada aturan Peraturan Badan Pusat Statistik Nomor 3 Tahun 2019 dengan tipe data teks.
* kategori\_disabilitas: menyatakan kategori dari jenis disabilitas dengan tipe data teks.
  + cacat fisik: menyatakan jenis disabilitas dengan cacat fisik.
  + cacat fisik dan mental: menyatakan jenis disabilitas dengan cacat fisik dan cacat mental.
  + cacat mental/jiwa: menyatakan jenis disabilitas dengan cacat mental/jiwa.
  + cacat netra/buta: menyatakan jenis disabilitas dengan cacat netra/buta.
  + cacat rungu/wicara: menyatakan jenis disabilitas dengan cacat rungu/wicara.
  + cacat lainnya: menyatakan jenis disabilitas dengan cacat rselain yang disebutkan diatas.
* jumlah\_penduduk: menyatakan jumlah penduduk penyandang disabilitas dengan tipe data numerik.
* satuan: menyatakan satuan dari pengukuran jumlah penduduk penyandang disabilitas dalam jiwa dengan tipe data teks.
* tahun: menyatakan tahun produksi data dengan tipe data numerik.



**Gambar 1. Dataset**

**3.2 Preprocessing**



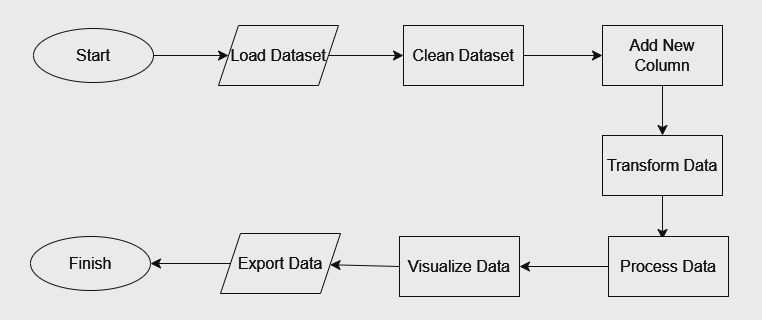
**Gambar 2. Flowchart preprocessing**

Skenario :

* Pertama-tama, dataset perlu di-load ke dalam Spark menggunakan salah satu API Spark, misalnya DataFrames atau Datasets. Dataset ini bisa berasal dari file CSV, JSON, atau sumber data lainnya.
* Selanjutnya, dataset harus di-clean dari data yang tidak valid dengan menggunakan API SparkSQL, DataFrames atau Datasets.
* Lalu manambahkan kolom baru yang akan diperlukan
* Setelah itu, data perlu di transformasikan untuk mempersiapkannya untuk analisis lebih lanjut.
* Selanjutnya, dataset diproses pada SparkSQL untuk divisualisasikan dan diolah sesuai kebutuhan.
* Hasil pengolahan data dapat diekspor ke dalam format yang berguna seperti CSV atau JSON untuk digunakan di aplikasi atau platform lain.

**3.3 Implementasi Spark Machine Learning**

Diagram flowchart Spark ML :

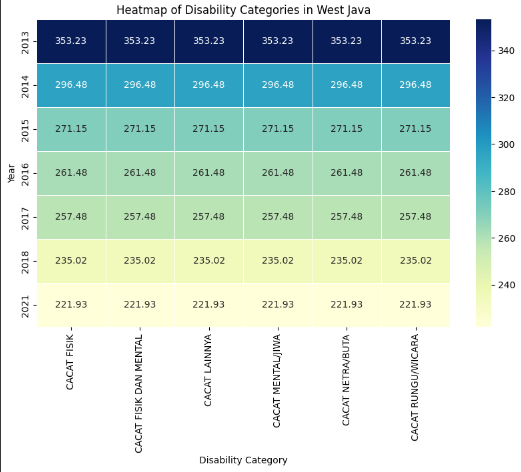


**Gambar 3. Flowchart Spark Machine Learning**

pada tahap ini dilakukan pengelompokan data menggunakan machine learning untuk dapat divisualisasikan berupa grafik.

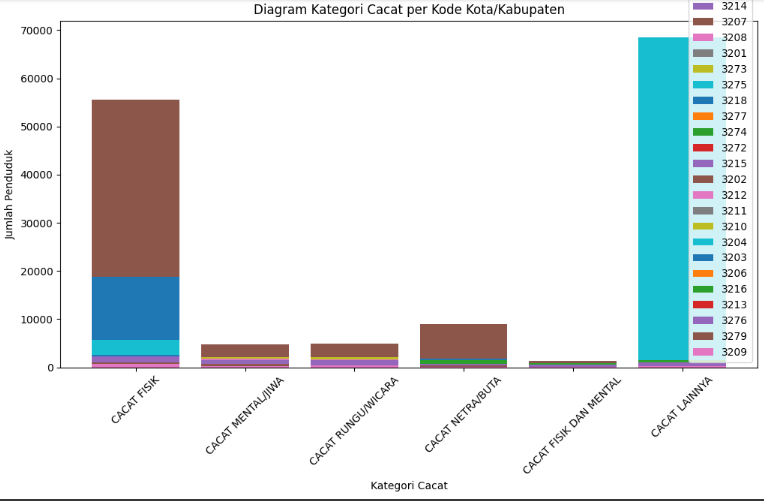
1. **Hasil dan Pembahasan (or Results and Analysis)**

Hasil dari pengolahan data dengan cara mengurutkan data berdasarkan tahun secara ascending lalu pivot tabel dengan menggunakan fungsi agregasi mean untuk menggabungkan nilai-nilai yang memiliki tahun yang sama dan hasilnya akan ditampilkan berupa diagram heatmap dibawah ini.

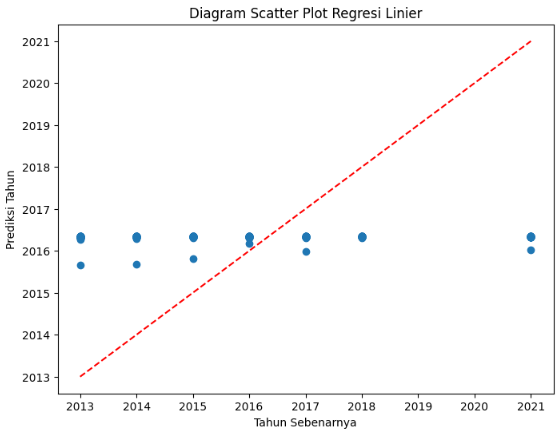


**Gambar 4. Heatmap Kategori Disablitas**

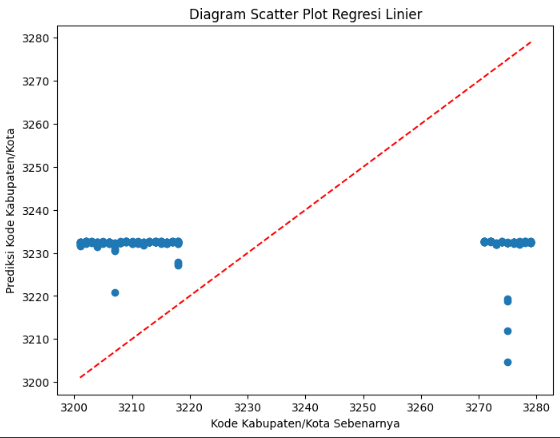
Selanjutnya adalah mengkategorikan setiap jenis cacat dan jumlahnya sesuai dengan kode kabupaten/kota. Pada diagram ini nantinya dapat diketahui jumlah cacat dari setiap kota/kabupaten berdasarkan jenis cacatnya.



**Gambar 5 Diagram Kategori Cacat per Kode Kota/Kabupaten**

****

**Gambar 6 Diagram Scatter Plot Regresi Linear (Tahun Sebenarnya)**

****

**Gambar 7 Diagram Scatter Plot Regresi Linear (kode kabupaten/kota sebenarnya)**

# Kesimpulan

Dari pembahasan di atas dengan pengelompokan data dari setiap kategori cacat dan dapat diketahui jumlah dari setiap kota/kabupaten diharapkan dapat memudahkan dalam mengatasi masalah aksesibilitas yang dihadapi oleh penyandang disabilitas. Paper ini menyimpulkan bahwa Big Data Analytics memiliki potensi besar dalam meningkatkan aksesibilitas bagi penyandang disabilitas. Dalam memanfaatkan kumpulan data yang ada, kita dapat mengidentifikasi tantangan, mengembangkan solusi yang efektif, dan memberikan rekomendasi kebijakan dan praktik terbaik. Diharapkan paper ini dapat memberikan sumbangan penting dalam mengatasi masalah aksesibilitas yang dihadapi oleh penyandang disabilitas, serta berkontribusi pada upaya menciptakan masyarakat yang lebih inklusif dan memberikan kesempatan setara bagi semua individu.

# Referensi (Reference)

[1] E. Santucci, L. Didaci, G. Fumera, and F. Roli, “A Parameter Randomization Approach for Constructing Classifier Ensembles,” *Pattern Recognition.*, vol. 69, pp. 1–13, 2017.

[2] P. Panov and S. Džeroski, “Combining Bagging and Random Subspaces to Create Better Ensembles,” in *Proceedings of the 7th International Conference on Intelligent Data Analysis*, 2007, pp. 118–129.

[3] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms: Second Edition*. New Jersey: John Wiley and Sons, 2004.

[4] R. M. Barts, “The Stub Loaded Helix: A Reduced Size Helical Antenna,” *Doctoral Dissertation*, 2003. [Online]. Available: http://hdl.handle.net/10919/29728.