Analisis Big Data UMKM Menggunakan Apache Hadoop-Hive dengan Metode *Forecasting Time Series*

Putu Denanta Bayuguna Perteka¹, Nyoman Gunantara², Ida Bagus Alit Swamardika³ [Submission: 05-11-2024, Accepted: 14-12-2024]

Abstract— MSMEs (Micro, Small, and Medium Enterprises) play a vital role in economic growth, creating jobs, and improving people's welfare. However, MSMEs often face challenges in managing sales data and performance analysis, resulting in many businesses closing down every year. The Covid-19 pandemic has worsened this situation, with many MSMEs experiencing a decline in revenue and product demand. In the past five years, the application of big data analysis has increased rapidly, helping with fast and accurate decisionmaking. This technology, especially the time series forecasting method, allows predictions of future values based on historical data, thus supporting MSMEs in increasing productivity and competitiveness. The industrial revolution 4.0 brings big data and blockchain-based business intelligence that accelerates the supply chain and trade. This helps the government in setting regulations and making specific decisions for MSMEs. Although there are challenges in building infrastructure and deploying this technology, the benefits are enormous, such as understanding community trends through sentiment analysis on social media, improving company image, planning businesses based on consumer behavior, and making fast and accurate decisions. Apache Hadoop and Hive are efficient tools for managing and analyzing big data. In this study, MSME data in Bali Province was analyzed using the simple moving average forecasting time series method. The results of this analysis are used as a basis for recommendations and improving management, and are expected to be a reference for better decision making.

Intisari— UMKM (Usaha Mikro, Kecil, dan Menengah) memainkan peran penting dalam pertumbuhan ekonomi, menciptakan lapangan kerja, dan meningkatkan kesejahteraan masyarakat. Namun, UMKM sering menghadapi tantangan dalam pengelolaan data penjualan dan analisis kinerja, yang mengakibatkan banyaknya usaha yang tutup setiap tahun. Pandemi Covid-19 memperburuk situasi ini, dengan banyak UMKM mengalami penurunan pendapatan dan permintaan produk. Dalam lima tahun terakhir, penerapan analisis big data telah meningkat pesat, membantu pengambilan keputusan yang cepat dan akurat.

Teknologi ini, terutama metode forecasting time series, memungkinkan prediksi nilai masa depan berdasarkan data historis, sehingga dapat mendukung UMKM dalam

¹Mahasiswa, Program Studi Magister Teknik Elektro Universitas Udayana, Jalan P.B. Sudirman, Dangin Puri Klod, Kec. Denpasar Bar., Kota Denpasar, Bali 80234 (tlp: 085931270707; e-mail: perteka.2281711005@student.unud.ac.id)

^{2,3} Dosen, Program Studi Magister Teknik Elektro Universitas Udayana, Jalan P.B. Sudirman, Dangin Puri Klod, Kec. Denpasar Bar., Kota Denpasar, Bali 80234 (telp: 081558140279; 085237256257; e-mail: gunantara@unud.ac.id;

Putu Denanta Bayuguna Perteka: Analisis Big Data UMKM...

gusalit@unud.ac.id)

meningkatkan produktivitas dan daya saing. Revolusi industri 4.0 membawa kecerdasan bisnis berbasis big data dan blockchain yang mempercepat rantai pasokan dan perdagangan. Hal ini membantu pemerintah dalam menetapkan regulasi dan pengambilan keputusan spesifik untuk UMKM. Meski ada tantangan dalam membangun infrastruktur dan menyebarkan teknologi ini, manfaatnya sangat besar, seperti memahami tren masyarakat melalui analisis sentimen di media meningkatkan citra perusahaan, merencanakan berdasarkan perilaku konsumen, dan membuat keputusan cepat dan akurat. Apache Hadoop dan Hive adalah alat yang efisien untuk mengelola dan menganalisis big data. Dalam penelitian ini, data UMKM di Provinsi Bali dianalisis menggunakan metode forecasting time series simple moving average. Hasil analisis ini digunakan sebagai dasar untuk memberikan rekomendasi dan meningkatkan pengelolaan UMKM, diharapkan dapat menjadi acuan dalam pengambilan keputusan yang lebih baik.

Kata Kunci— UMKM; big data; apache hadoop; hive; forecasting time series; simple moving average.

I. PENDAHULUAN

UMKM (Usaha Mikro, Kecil, dan Menengah) memiliki peranan strategis dalam perekonomian, khususnya dalam penciptaan lapangan kerja dan peningkatan kesejahteraan masyarakat. Data dari Kementerian Koperasi dan UKM menunjukkan bahwa pada tahun 2019, terdapat 65,4 juta UMKM yang menyerap tenaga kerja sebesar 123,3 ribu orang dan menyumbang 60,5% dari PDB Nasional [1]Meskipun demikian, UMKM menghadapi tantangan signifikan dalam pengelolaan data penjualan dan analisis kinerja usaha. Setiap tahunnya, banyak UMKM yang terpaksa menutup usaha karena produk yang kurang diminati atau kurang tepat sasaran, mengakibatkan penurunan keuntungan hingga kebangkrutan.

Dampak pandemi Covid-19 memperburuk situasi ini. Berdasarkan survei terhadap 1.180 pelaku UMKM, 48% di antaranya mengalami kesulitan bahan baku, 77% melaporkan penurunan pendapatan, 88% mengalami penurunan permintaan produk, dan 97% melaporkan penurunan nilai aset (Kementerian Koordinator Bidang Perekonomian RI, 2022). Dalam lima tahun terakhir, penelitian terhadap penerapan analisis big data meningkat, menunjukkan efektivitasnya dalam mendukung pengambilan keputusan cepat dan akurat, memperkirakan permintaan rantai pasokan, menilai risiko, dan menyederhanakan rantai pasokan [2].

Analisis big data menjadi penting dalam meningkatkan produktivitas dan pengambilan keputusan cepat bagi UMKM. Salah satu metode yang digunakan adalah forecasting time series, yang memungkinkan prediksi nilai masa depan berdasarkan data historis. Di era revolusi industri 4.0,



pengembangan sistem kecerdasan bisnis berbasis big data dan blockchain sangat penting untuk meningkatkan daya saing global UMKM, sekaligus mempercepat arus rantai pasokan dan perdagangan. Hal ini dapat mendukung pemerintah dalam menetapkan regulasi, serta memfasilitasi pelatihan spesifik yang dibutuhkan UMKM. Data akurat dan real-time berperan dalam manajemen bisnis yang lebih baik, membantu perusahaan meningkatkan keunggulan kompetitif. Namun, tantangan terbesar dalam penerapan big data terletak pada pembangunan infrastruktur dan penyebaran teknologi melalui pelatihan [3].

Big data UMKM merupakan kumpulan data yang besar dari kelompok usaha mikro, kecil dan menengah, yang mencakup berbagai informasi seperti data transaksi, produktivitas, investasi dan tren produksi. Karakteristik utama dari big data UMKM adalah volumenya yang besar. Big data UMKM khususnya provinsi Bali memiliki nilai karakteristik yang unik karena terkait dengan budaya dan pariwisata daerah seperti kerajinan kain tenun, kerajinan dari kayu, bambu, dll

Big data memerlukan pengelolaan yang cermat, termasuk cara penyimpanan, pencarian, distribusi, hingga analisis data. Jika diterapkan dengan baik, big data dapat memberikan berbagai manfaat bagi UMKM, seperti (1) memahami tren dan kebutuhan masyarakat melalui analisis sentimen di media sosial, (2) meningkatkan citra perusahaan, (3) perencanaan usaha melalui analisis perilaku konsumen, dan (4) mendukung pengambilan keputusan cepat dan akurat berbasis data [4].

Apache Hadoop adalah salah satu tools yang digunakan untuk mengelola dan menganalisis big data dengan cepat dan efisien, menyediakan infrastruktur database dan distribusi data yang skalabel [5]. Hive, yang dibangun di atas Apache Hadoop, berfungsi untuk menyimpan dan memproses set data besar secara efisien dengan memungkinkan pengguna mengakses data petabyte menggunakan SQL.

Data UMKM dari Provinsi Bali dianalisis menggunakan metode forecasting time series berbasis moving average, yang memanfaatkan nilai rata-rata dari sejumlah observasi untuk memprediksi nilai masa depan berdasarkan data historis dalam rentang waktu tertentu. Dataset ini mencakup beberapa kategori Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) sebagai acuan dalam memberikan rekomendasi kinerja per kategori UMKM. Hasil analisis ini diharapkan dapat digunakan sebagai dasar pengambilan keputusan yang lebih baik dalam pengelolaan dan peningkatan kinerja UMKM.

Penelitian ini menggabungkan dan mengembangkan beberapa penelitian sebelumnya tentang analisis big data menggunakan metode forecasting time series, yang diimplementasikan dengan Apache Hadoop-Hive. Tujuan utamanya adalah meningkatkan produktivitas UMKM serta mendukung pemerintah dan pelaku usaha dalam pengambilan keputusan berdasarkan data yang akurat dan up-to-date.

II. STUDI PUSTAKA

Penelitian terkait penerapan big data dan forecasting pada UMKM telah banyak dilakukan sebelumnya:

1) **Chatterjee et al. (2023)** mengungkapkan bahwa teknologi analisis big data dapat mendukung proses

- pengambilan keputusan yang cepat dan akurat, terutama dalam memperkirakan permintaan rantai pasokan [2].
- 2) **Wibawa et al. (2022)** menunjukkan bahwa penggunaan Apache Spark lebih cepat dibandingkan Hadoop MapReduce dalam pengolahan data teks [6].
- 3) Lusiana dan Yuliarty (2020) membandingkan beberapa metode forecasting dan menemukan bahwa metode Exponential Smoothing dengan parameter tertentu menghasilkan tingkat kesalahan yang paling kecil [7].
- 4) Nurul Azizah dan Henry Saptono (2020) melakukan pen performa RDBMS Mysql dan Apache Hive-Hadoop menunjukan bahwa performa dari Apache Hive Hadoop memiliki kecepatan yang siginifikan tinggi dibandingkan dengan RDBMS Mysql, hal ini dikarenakan Hadoop merupakan file system yang terdistribusi sedangkan Mysql merupakan database untuk penyimpanan data. [8]
 - M Rifqi, M Riswandha dan Imam Ghozali (2024) membandingkan dua komponen utama ekosistem Hadoop, yaitu Hive dan HBase, melalui tinjauan literatur sistematis. HBase menggunakan model wide-column store yang unggul dalam operasi baca/tulis dengan latensi rendah dan cocok untuk pembaruan data realtime. Sebaliknya, Hive mengadopsi model relasional dengan antarmuka SQL (HiveQL), dirancang untuk analisis data besar dengan memanfaatkan model MapReduce. Meskipun HBase lebih baik untuk operasi dengan kebutuhan latensi rendah, Hive unggul dalam eksekusi kueri data besar. Dari segi keamanan, HBase menawarkan enkripsi dan algoritma masking untuk melindungi data sensitif, sementara Hive mendukung autentikasi, otorisasi, dan enkripsi untuk meniaga kerahasiaan data. Penelitian ini menyimpulkan bahwa pemilihan antara Hive dan HBase bergantung pada kebutuhan spesifik aplikasi, seperti analisis data skala besar atau akses data cepat.[9]
- Ayu Sintiawati dan Putu Purbadharmaja (2024) menganalisis UMKM di Kota Denpasar yang memegang peranan penting dalam perekonomian daerah sebagai penyedia lapangan kerja dan pendorong pertumbuhan ekonomi lokal. Dalam penelitian ini menunjukkan bahwa faktor tenaga kerja, modal, harga, dan teknologi secara signifikan memengaruhi pendapatan UMKM di wilayah ini. Tenaga kerja yang terampil, ketersediaan modal yang memadai, penentuan harga yang kompetitif, serta pemanfaatan teknologi seperti aplikasi SIPLah. Aplikasi SIPLah membantu pelaku UMKM memperluas akses pasar, meningkatkan pemasaran, dan mempermudah proses transaksi, sehingga berkontribusi pada pendapatan yang lebih tinggi. Dengan didukung oleh tenaga kerja yang terampil, modal yang cukup, serta harga yang kompetitif, adopsi teknologi seperti SIPLah menjadi solusi strategis bagi UMKM Denpasar untuk tetap bertahan dan berkembang [10]

Dalam penelitian ini, menggunakan Apache Hadoop sebagai tempat penyimpanan data secara terdistribusi ,Hive digunakan sebagai pengelolaan data dengan HiveQL yang memiliki performa cepat dalam menkueri data besar, dan metode forecasting time series yang digunakan adalah *Simple*

Moving Average, metode ini dipilih karena kesederhanaannya serta kemampuannya dalam menangani data deret waktu.

A. Analisis Big Data

Big data adalah kumpulan data digital dalam volume besar, terdiri dari data terstruktur, semi-terstruktur, dan tidak terstruktur, yang dihasilkan dan dikumpulkan oleh organisasi untuk diolah menjadi informasi berharga guna mendukung pengambilan keputusan. Analisis big data bertujuan menemukan pola dari data yang beragam jenis dan sumbernya serta berukuran besar dengan menggunakan metode komputasi sistematis. Keunggulan big data meliputi optimalisasi biaya, peningkatan efektivitas dan efisiensi, serta kemampuan memprediksi perilaku pelanggan dan permintaan pasar. Pengelolaan big data mencakup tiga tahapan utama: integrasi data, manajemen data, dan analisis. Tahap integrasi mengumpulkan data dari berbagai sumber seperti transaksi data pelanggan. Manaiemen data melibatkan pengelompokan dan penyimpanan data dalam sistem manaiemen basis data (DBMS). Tahap analisis mengeksplorasi dan memproses data untuk menghasilkan informasi yang relevan [11]

B. Apache Hadoop

Apache Hadoop adalah kerangka kerja yang dirancang untuk menyimpan dan memproses data dalam volume besar secara terdistribusi dan paralel dalam kluster komputer. Hadoop memiliki beberapa fitur utama, yaitu Hadoop Distributed File System (HDFS), yang berfungsi sebagai penyimpanan terdistribusi dengan membagi data menjadi blok kecil dan mendistribusikannya ke berbagai node dalam kluster. Fitur lain adalah MapReduce, model pemrograman yang memproses data dalam dua tahap, yaitu Map untuk pemrosesan paralel dan Reduce untuk menggabungkan hasil pemrosesan. Selain itu, Hadoop dilengkapi dengan Yet Another Resource Negotiator (YARN), manajer sumber daya yang bertanggung jawab mengelola dan menjadwalkan aplikasi di dalam kluster, seperti MapReduce dan Spark. Hadoop Common juga disertakan sebagai utilitas umum yang menyediakan pengaturan dan alat bantu lainnya. Ekosistem Hadoop mencakup berbagai proyek tambahan seperti Apache Hive untuk query dan analisis data, Apache Pig untuk pemrosesan data, Apache HBase untuk penyimpanan data terdistribusi, serta Apache Spark untuk pemrosesan data cepat, yang semuanya mendukung kinerja Hadoop keseluruhan.

C. Hive

Hive adalah kerangka kerja data warehouse open source yang dibangun di atas Apache Hadoop. Hive digunakan untuk menjalankan query dan mengelola kumpulan data terdistribusi dengan pendekatan mirip SQL. Hive memudahkan dalam membentuk konsep data terstruktur dan menyediakan bahasa query yang dikenal sebagai HiveSQ.

D. Forecasting

Putu Denanta Bayuguna Perteka: Analisis Big Data UMKM...

Forecasting atau peramalan adalah proses dalam memperkirakan berapa kebutuhan dimasa mendatang yang meliputi kebutuhan dalam ukuran kuantitas, kualitas, waktu dan lokasi yang dibutuhkan dalam rangka memenuhi permintaan barang maupun jasa [12].

E. Simple Moving Average

SMA adalah metode yang paling sederhana di mana setiap data dalam periode yang digunakan memiliki bobot yang sama. Rata-rata dihitung dengan menjumlahkan nilai-nilai dari sejumlah periode tertentu dan membaginya dengan jumlah periode tersebut.

$$SMAt = (Xt + Xt - 1 + Xt - 2 + Xt - 3 + \dots + Xt - (n-1))/n$$
 (1)

Keterangan:

SMAt : Simple moving average pada periode ke-t.

Xt : Nilai data pada periode ke-t.

n : Jumlah periode yang digunakan untuk menghitung rata-rata.

F. UMKM

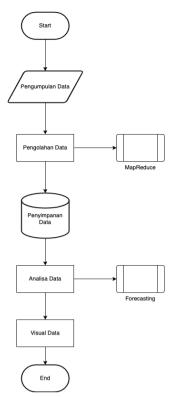
UMKM (Usaha Mikro, Kecil, dan Menengah) dikategorikan berdasarkan ukuran operasionalnya. Usaha Mikro memiliki jumlah karyawan yang relatif sedikit, dikelola oleh perorangan atau keluarga, dan memiliki skala operasional kecil. Usaha Kecil memiliki skala operasional dan jumlah karyawan yang lebih besar dibandingkan Usaha Mikro, dengan omzet yang juga lebih tinggi. Sementara itu, Usaha Menengah lebih besar dari Usaha Kecil tetapi lebih kecil dari usaha besar atau koperasi, memiliki struktur organisasi yang jelas dan skala operasional yang lebih luas.

III. METODOLOGI

Penelitian menggunakan pendekatan kuantitatif dengan teknik pengumpulan data dari database UMKM Provinsi Bali. Data diproses menggunakan Apache Hadoop-Hive, yang menyimpan data di Hadoop Distributed File System (HDFS). Metode forecasting time series yang diterapkan adalah Simple Moving Average untuk menganalisis tren dan memprediksi kinerja masa depan UMKM. Visualisasi hasil analisis dilakukan menggunakan Matplotlib untuk memberikan gambaran yang lebih jelas.

A. Rancangan Penelitian





Gambar 1: Rancangan Penelitan

Penelitian ini akan dilakukan dengan langkah-langkah berikut:

1) Pengumpulan Data

Data akan dikumpulkan dari berbagai sumber seperti database dan file excel yang memuat informasi terkait kinerja UMKM di Provinsi Bali selama periode yang ditentukan.

2) Pengolahan Data

Data yang telah terkumpul akan diproses menggunakan metode MapReduce melalui Apache Hadoop-Hive. MapReduce adalah model pemrograman yang memungkinkan pemrosesan data besar dengan membaginya menjadi bagianbagian kecil yang dapat diolah secara paralel.

3) Penyimpanan Data

Data hasil olahan akan disimpan dalam Hadoop Distributed File System (HDFS), yang memungkinkan penyimpanan data dalam skala besar dan memastikan akses yang cepat dan efisien.

4) Analisis Data

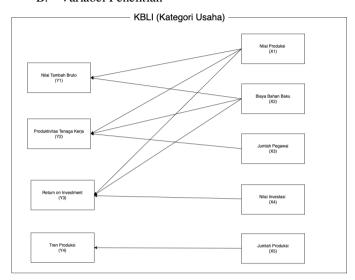
Data yang disimpan kemudian akan dianalisis menggunakan metode *forecasting time series*, khususnya metode *Simple Moving Average*. Analisis ini bertujuan untuk memprediksi kinerja UMKM di masa depan berdasarkan data historis.

5) Visualisasi Data

Hasil analisis akan divisualisasikan menggunakan Tableau. Tableau adalah alat visualisasi data yang memungkinkan presentasi data dalam bentuk grafik dan diagram yang mudah dipahami, sehingga membantu dalam interpretasi hasil analisis dan mendukung pengambilan keputusan yang lebih baik.

Dengan langkah-langkah ini, penelitian diharapkan dapat memberikan wawasan mendalam mengenai kinerja UMKM di Provinsi Bali dan membantu meningkatkan pengelolaan serta produktivitas mereka melalui penerapan teknologi big data dan metode forecasting yang tepat.

B. Variabel Penelitian



Gambar 2: Variabel Penelitian

Pada penelitian ini terdapat 5 variabel bebas (x) yaitu (x_1) nilai produksi, (x_2) biaya bahan baku, (x_3) jumlah pegawai, (x_4) nilai investasi, (x_5) jumlah produksi dan 1 variabel kontrol yaitu kategori KBLI yang dapat menghasilkan berbagai macam kondisi (y) seperti nilai tambah bruto (y_1) , produktivitas tenaga kerja (y_2) , return on investment (y_3) dan tren produksi (y_4) . nilai tambah bruto (y_1) dipengaruhi oleh nilai produksi (x_1) dan biaya bahan baku (x_2) . sedangkan untuk produktivitas tenaga kerja (y_2) dipengaruhi oleh nilai tambah bruto (y_1) dan jumlah pegawai (x_3) . return on investment (y_3) bergantung pada nilai tambah bruto (y_1) dan nilai investasi (x_4) , dan untuk tren produksi dipengaruhi oleh jumlah produksi pada periode sebelumnya.

C. Dataset

Dataset yang digunakan dalam penelitian ini mencakup data UMKM di Provinsi Bali selama periode 2018 hingga 2023. Data tersebut berasal dari 9 kabupaten/kota, yaitu Badung, Denpasar, Gianyar, Klungkung, Karangasem, Buleleng, Bangli, Tabanan, dan Jembrana. Dataset terdiri dari nama perusahaan, pemilik, alamat, desa, kecamatan, kabupaten, bentuk badan usaha, tahun izin, KBLI, produk, tenaga kerja, nilai investasi, jumlah produksi, nilai produksi, nilai BB/BP dan tahun pendataan. Dataset ini akan diproses menggunakan Hive dan disimpan dalam Hadoop Distributed File System (HDFS). Data yang diuji mencakup 146 kategori KBLI, dengan total jumlah data yang tersedia sebanyak 71.400 dengan besaran file 1977Kb dan akan bertambah setiap tahunnya.

Majalah Ilmiah Teknologi Elektro, Vol.23, No.2, Juli-Desember 2024 DOI: https://doi.org/10.24843/MITE.2024.v23i02.P10

D. Apache Hadoop File System (HDFS)

Apache Hadoop File System atau HDFS merupakan sistem penyimpanan terdistribusi yang dirancang unutk menyimpan dan mengelola data dalam jumlah besar untuk menangani volume data yang sangat besar dengan efisiensi tinggi, skalabilitas dan tolerasi terhadap gangguan.

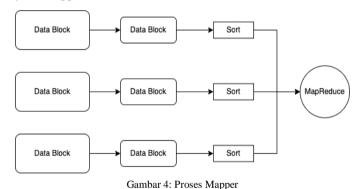
```
root@8673397ab88:/employee# hadoop fs -put DATA_UMMM_2019.csv hdfs://namenode:8020/user/hive/warehouse/umkm.db/pendataan_umkm
rout@8673937ab88:/employee# hadoop fs -put DATA_UMMM_2028.csv hdfs://namenode:8020/user/hive/warehouse/umkm.db/pendataan_umkm
rout@8673937ab88:/employee# rout@8673937ab88:/employee# hadoop fs -put DATA_UMMM_2021.csv hdfs://namenode:8020/user/hive/warehouse/umkm.db/pendataan_umkm
rout@86739397ab88:/employee# rout@8739397ab88:/employee# hadoop fs -put DATA_UMMM_2021.csv hdfs://namenode:8020/user/hive/warehouse/umkm.db/pendataan_umkm
rout@86739397ab88:/employee# rout@873997ab88:/employee# hadoop fs -put DATA_UMMM_2022.csv hdfs://namenode:8020/user/hive/warehouse/umkm.db/pendataan_umkm
rout@86739397ab88:/employee# hadoop fs -put DATA_UMMM_2023.csv hdfs://namenode:8020/user/hive/warehouse/umkm.db/pendataan_umkm
```

Gambar 3: Memindahkan file ke dalam HDFS

hadoop fs -put untuk memindahkan file dari sistem lokal ke dalam Hadoop Distributed File System (HDFS). Dalam hal ini, file-file CSV yang berisi data UMKM dari tahun 2019 hingga 2023 diunggah ke direktori HDFS pada jalur /user/hive/warehouse/umkm.db/pendataan_umkm. Proses ini bertujuan untuk menyimpan data secara terdistribusi dalam kluster Hadoop, memungkinkan akses dan pemrosesan data besar secara paralel dan efisien. Jalur tujuan di HDFS mungkin diatur sebagai basis data dalam Hive untuk mempermudah analisis data lebih lanjut.

E. MapReduce

MapReduce merupakan model yang digunakan untuk memproses data yang besar secara terdistribusi dan paralel. Pemrosesan data dalam MapReduce dibagi menjadi 2 tahap yaitu mapper dan reducer.



Pada tahap ini, data besar yang tersimpan dalam bentuk Data Block dibagi menjadi bagian-bagian kecil dan didistribusikan ke node-node dalam kluster Hadoop untuk diproses. Fungsi Mapper memetakan data menjadi pasangan kunci-nilai (key-value pairs) yang sesuai untuk analisis lebih lanjut. Setelah pemetaan, data disortir berdasarkan kunci untuk mempersiapkannya masuk ke tahap Reducer. Proses ini

untuk mempersiapkannya masuk ke tahap Reducer. Proses ini memungkinkan Hadoop mengolah data dalam skala besar secara paralel, meningkatkan efisiensi dan kecepatan pemrosesan.

Putu Denanta Bayuguna Perteka: Analisis Big Data UMKM...

MapReduce Reduce File

Gambar 5: Proses Reducer

Gambar tersebut menunjukkan proses Reducer yang dilakukan setelah tahap Mapper dalam model MapReduce. Pada tahap ini, data yang telah dikelompokkan dihitung total kemunculannya, dan hasilnya disimpan ke dalam sebuah file. File ini berisi informasi akhir dari data yang telah diolah, seperti jumlah kemunculan kata atau data lainnya. Tahap Reducer merupakan bagian penting karena mengubah data yang dikelompokkan menjadi informasi yang bermakna dan dapat digunakan.

```
Notes in Secret overmeits table part partition(041)

Select though production, and series, smaller produkts, nilai_needuks, nilai_bo_bp, kbli from pen datam_namber;

Select though production production products, nilai_needuks, nilai_needuks, nilai_bo_bp, kbli from pen datam_namber;

2024-04-17799:31:57,593 TAFO [main] org.apache.hadoop.hive.comf.HiveComf - Using the default value passed in fer log id: 2cf7ab39-5d15-4Mbd-916d-frea55ef1b602 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist. htms://localhost/sid=4Mbd-916d-frea55ef1b602 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist. htms://localhost/sid=4Mbd-916d-frea55ef1b602 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist. htms://localhost/sid=4Mbd-916d-frea55ef1b602 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist. htms://localhost/sid=4Mbd-916d-frea55ef1b602 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist. htms://localhost/sid=4Mbd-916d-frea55ef1b602 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist. htms://localhost/sid=4Mbd-916d-frea55ef1b602 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist. htms://localhost/sid=4Mbd-916d-frea55ef1b602 main] org.apache.hadoop.hive.2024-04-17_69-11-52_909_1338678V11061141972-1/Degry ID - Admini-gradef109 of 100 tot of 3

Namehre of reduce tasks no specified. Estimated from input data size: 1

Namehre of reduce tasks no specified. Estimated from input data size: 1

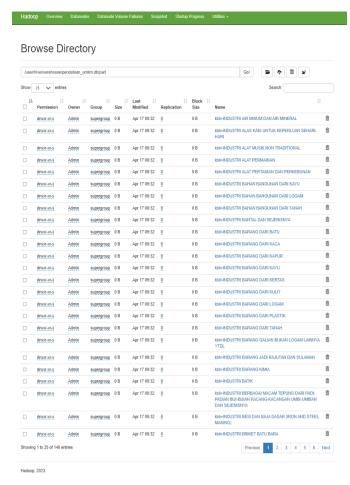
Size shive.exc.exc.evdoers.hytes_producers_common_sec_evdoers_nytes_producers_common_sec_evdoers_nytes_producers_common_sec_evdoers_nytes_producers_common_sec_evdoers_nytes_producers_common_sec_evdoers_nytes_producers_common_sec_evdoers_nytes_producers_common_sec_evdoers_nytes_producers_common_sec_evdoers_nytes_producers_common_sec_evdoers_nytes_producers_common_sec_evdoers_nytes_producers_common_sec_ev
```

Gambar 6: Map Reduce menggunakan Hadoop dan Hive SQL

Query HiveSQL tersebut digunakan untuk mengelola data UMKM menggunakan proses MapReduce dan menyimpannya dalam tabel partisi. Hasil dari pengolahan data ini akan disimpan di dalam Hadoop Distributed File System (HDFS) sebagai direktori data. Cluster dan Nodes

Gambar 7 menampilkan direktori Hadoop di dalam folder pendataan_umkm.db yang berisi partisi-partisi data. Setiap partisi merepresentasikan data yang diorganisir berdasarkan kategori tertentu, seperti KBLI industri air minum dan air mineral, KBLI industri alas kaki, dan lainnya. Direktori ini berfungsi sebagai lokasi penyimpanan dan pengorganisasian data yang nantinya akan dianalisis. Sistem Hadoop digunakan untuk mengelola data yang terkait dengan berbagai sektor industri yang tercantum dalam kode KBLI (Klasifikasi Baku Lapangan Usaha Indonesia). Terdapat 146 entri KBLI, menunjukkan keberagaman data dan mencakup banyak sektor industri di Provinsi Bali yang siap untuk dianalisis menggunakan Hive.





Gambar 7: Direktori HDFS

IV. HASIL DAN PEMBAHASAN

Pengujian dilakukan terhadap data yang berkaitan dengan industri kain tenun dengan menggunakan metode forecasting simple moving average. Pengujian dilakukan dengan menganalisis kegiatan-kegitan utama dalam industri, seperti produktivitas tenaga kerja, penjualan, pendapatan, dan tren produksi. Hasil dari pengujian diharapkan dapat memberikan wawasan baru terkait pengembangan industry kain tenun di Provinsi Bali, serta memberikan rekomendasi bagi pihak terkait dalam proses pengambilan keputusan.

TABEL 1 DATASET UMKM INDUSTRI KAIN TENUN

Indikator	2019	2020	2021	2022	2023
Nilai Tambah Bruto	2.434.82 5	2.510.66 5	2.413.95 4	1.837.44 5	1.980.59 5
Produktivita s Tenaga Kerja	252,94	255,88	265,625	265,625	264,063
Return on Investment	1,8835	1,98124	1,8690	1,225	1,3793
Tren Produksi	8.600	8.700	8.600	8.500	8.450

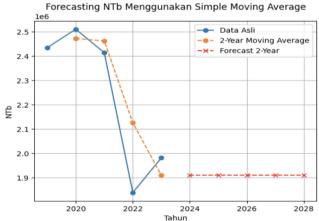
Tabel diatas merupakan data histori terkait dengan Data KBLI Industri kain tenun yang meliputi 4 indikator seperti Nilai Tambah Bruto (NTb), Produktivitas Tenaga Kerja, Return on Investment (ROI), dan Tren Produksi.

TABEL 2 HASIL PERAMALAN MENGGUNAKAN METODE SIMPLE MOVING AVERAGE

Forecasting							
2024	2025	2026	2027	2028	RMSE		
1.909.020	1.909.020	1.909.020	1.909.020	1.909.020	71.575		
264,844	264,844	264,844	264,844	264,844	9,12		
1,302	1,302	1,302	1,302	1,302	0,53		
8475	8475	8475	8475	8475	148,5		

Berdasarkan hasil peramalan, terdapat beberapa indikator utama yang diproyeksikan untuk periode 2024 hingga 2028. Nilai Tambah Bruto diperkirakan tetap stabil pada angka 1.909.020 setiap tahunnya dengan tingkat kesalahan (RMSE) sebesar 71.575. Produktivitas Tenaga Kerja juga menunjukkan kestabilan pada nilai 264,844 di seluruh periode yang sama, dengan tingkat kesalahan sebesar 9,12. Indikator Return on Investment (ROI) diprediksi konsisten pada angka 1,302 setiap tahunnya, dengan RMSE sebesar 0,53. Sementara itu, Tren Produksi diestimasikan akan tetap pada angka 8475 sepanjang periode peramalan dengan tingkat kesalahan sebesar 148,5. Peramalan ini mencerminkan stabilitas yang tinggi dalam berbagai aspek ekonomi, namun tetap memerlukan perhatian lebih untuk memastikan bahwa prediksi ini sesuai dengan kondisi nyata yang mungkin berubah seiring waktu.

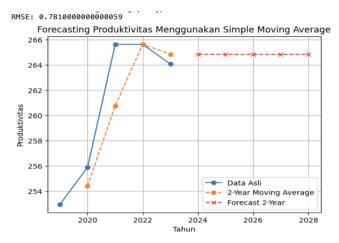
RMSE: 71575.0



Gambar 8: Grafik Forecasting Nilai Tambah Bruto (NTb)

Grafik tersebut memproyeksikan Nilai Tambah Bruto (NTb) menggunakan metode Simple Moving Average (SMA) dengan periode 2 tahun. Data asli dari 2019 hingga 2023 ditampilkan dengan garis biru, sementara garis oranye putusputus menunjukkan rata-rata bergerak 2 tahun. NTb mencapai puncak sekitar 2,5 juta pada 2020, namun turun di bawah 2 juta pada 2022, sebelum naik sedikit di 2023. Proyeksi dari 2024 hingga 2028, ditampilkan dengan garis merah putusputus, memperkirakan NTb stabil di sekitar 1,9 juta tanpa perubahan berarti. Nilai RMSE 71.575 menunjukkan

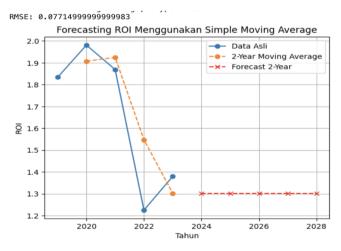
kesalahan yang cukup tinggi, menandakan metode ini kurang sensitif terhadap perubahan besar dalam industri.



Gambar 9: Grafik Forecasting Produktivitas

Grafik ini memproyeksikan produktivitas tenaga kerja menggunakan metode Simple Moving Average (SMA) 2 tahun. Data historis (garis biru) dari 2019 hingga 2023 menunjukkan peningkatan produktivitas dari 254 hingga puncak 266 pada 2021, sebelum sedikit menurun pada 2023. Garis oranye putus-putus menunjukkan rata-rata bergerak yang mengikuti pola serupa dengan perataan.

Proyeksi untuk 2024-2028 (garis merah putus-putus) memperkirakan produktivitas stabil di sekitar 265 tanpa fluktuasi besar. Nilai RMSE sebesar 0,781 menandakan kesalahan prediksi yang sangat kecil, menunjukkan akurasi tinggi meskipun kurang responsif terhadap perubahan atau tren baru di masa depan.

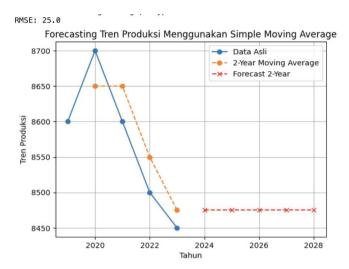


Gambar 10: Grafik Forecasting ROI

Grafik tersebut memperlihatkan peramalan Return on Investment (ROI) menggunakan metode Simple Moving Average (SMA) dengan periode 2 tahun. Data historis dari 2019 hingga 2023 (garis biru) menunjukkan ROI mencapai puncak 2,0 pada 2020 dan menurun drastis hingga 2022, di Putu Denanta Bayuguna Perteka: Analisis Big Data UMKM...

bawah 1,3, lalu sedikit naik pada 2023. Garis oranye putusputus menunjukkan rata-rata bergerak 2 tahun, sedangkan garis merah putus-putus menampilkan proyeksi ROI dari 2024 hingga 2028 yang stabil di sekitar 1,3.

Nilai RMSE sebesar 0,077 mengindikasikan tingkat kesalahan prediksi yang sangat kecil, menunjukkan akurasi tinggi berdasarkan data historis. Namun, karena sifat rata-rata bergerak, metode ini mungkin kurang responsif terhadap perubahan besar di masa depan.



Gambar 11: Grafik Forecasting Tren Produksi

Grafik ini memperlihatkan peramalan tren produksi dengan metode Simple Moving Average (SMA) 2 tahun. Data historis dari 2019 hingga 2023 (garis biru) menunjukkan produksi mencapai puncak 8.700 unit pada 2020, lalu menurun hingga 8.450 unit pada 2023. Garis oranye putus-putus menampilkan rata-rata bergerak yang mengikuti penurunan tersebut. Proyeksi untuk 2024 hingga 2028 (garis merah putus-putus) memperkirakan produksi akan stabil di sekitar 8.450 unit.

Nilai RMSE sebesar 25 mengindikasikan tingkat kesalahan yang relatif kecil, menunjukkan prediksi yang cukup akurat. Namun, metode SMA cenderung meratakan fluktuasi, sehingga mungkin kurang responsif terhadap perubahan signifikan di masa depan.

V. KESIMPULAN

penelitian ini menunjukkan Berdasarkan bahwa implementasi teknologi big data pada UMKM di Provinsi Bali menggunakan Apache Hadoop dan Hive mampu meningkatkan efisiensi dalam penyimpanan dan pemrosesan data, serta memberikan prediksi akurat terhadap indikator ekonomi seperti Nilai Tambah Bruto (NTb), Produktivitas Tenaga Kerja, Return on Investment (ROI), dan Tren Produksi, meskipun metode Simple Moving Average (SMA) yang digunakan kurang responsif terhadap perubahan besar dalam data historis. Dalam hal ini pemerintah Provinsi Bali memiliki peranan penting dalam memperkuat infrastruktur teknologi memberikan pelatihan kepada **UMKM**



memaksimalkan penggunaan big data dalam pengambilan keputusan. Penggunaan metode forecasting yang lebih canggih, seperti ARIMA atau Machine Learning, serta integrasi data real-time, juga perlu dipertimbangkan untuk meningkatkan akurasi dan relevansi prediksi di masa depan. Untuk memanfaatkan hasil penelitian ini secara optimal, beberapa rekomendasi praktis dapat diajukan.

1. UMKM:

- Mengadopsi teknologi big data dalam operasional sehari-hari, seperti pemantauan stok, analisis tren pasar, dan perencanaan produksi.
- Memanfaatkan pelatihan dan fasilitas teknologi yang disediakan pemerintah untuk meningkatkan literasi digital dan kemampuan analisis data.
- Mengintegrasikan metode forecasting yang lebih canggih, seperti ARIMA atau algoritma Machine Learning, guna meningkatkan akurasi prediksi dan responsivitas terhadap dinamika pasar.

Pemerintah:

- Menyediakan infrastruktur teknologi yang andal, seperti jaringan internet cepat dan akses perangkat lunak big data, khususnya di wilayah pedesaan.
- Mengembangkan kebijakan yang mendorong kolaborasi antara UMKM dan sektor swasta dalam adopsi teknologi big data.
- Memberikan subsidi atau insentif untuk pelaku UMKM yang aktif mengadopsi dan memanfaatkan teknologi big data.
- Mendorong pembaruan data secara berkala dengan membangun platform integrasi data real-time yang dapat diakses oleh UMKM.

Kolaborasi dengan Sektor Swasta:

- Melibatkan perusahaan teknologi dalam pengembangan aplikasi atau alat yang terjangkau dan sesuai dengan kebutuhan UMKM.
- Memanfaatkan keahlian sektor swasta pengembangan model prediksi berbasis big data yang lebih akurat dan relevan.diberikan relevan dan efektif dalam mendukung pertumbuhan dan keberlanjutan UMKM di Bali.

REFERENSI

- [1] Fajarharapan, "Kontribusi UMKM dalam Perekonomian Indonesia," https://djpb.kemenkeu.go.id/kppn/lubuksikaping/id/datapublikasi/artikel/3134-kontribusi-umkm-dalam-perekonomianindonesia.html, Pasaman, Jun. 26, 2023.
- S. Chatterjee, R. Chaudhuri, S. Gupta, U. Sivarajah, and S. Bag, [2] "Assessing the impact of big data analytics on decision-making processes, forecasting, and performance of a firm," Technol 196, Nov. Forecast Soc Change, vol. 2023. 10.1016/j.techfore.2023.122824.
- [3] M. Zaky, H. Jalan, T. Ryacudu, and L. Selatan, "Peluang Implementasi Teknologi Big Data Dan Block Chain Untuk

- Peningkatan Kinerja Perdagangan Pada Sektor Umkm Di Indonesia Pada Era Industri 4.0," 2019.
- ۲**4**1 A. Oktatriani, C. Destyana Putri, and D. Terttiaavini, "Peran Analisis Big Data Dalam Sektor Industri Di Indonesia," Jurnal Nasional Komputasi dan Teknologi Informasi, vol. 6, no. 3, 2023.
- [5] S. Oliviandi, A. B. Osmond, and R. Latuconsina, "Implementasi Apache Spark Pada Big Data Berbasis Hadoop Distributed File System," 2018.
- C. Wibawa, S. Wirawan, M. Mustikasari, and D. T. Anggraeni, [6] "Komparasi Kecepatan Hadoop Mapreduce Dan Apache Spark Dalam Mengolah Data Teks," Jurnal Ilmiah MATRIK, vol. 24, no. 1,
- [7] A. Lusiana and P. Yuliarty, "PENERAPAN METODE PERAMALAN (FORECASTING) PADA PERMINTAAN ATAP di PT X." 2020.
- N. Azizah and H. Saptono, "Uji Performa dan Perbandingan [8] RDBMS Mysql Dan Hive-Hadoop," Jurnal Informatika Terpadu, 20–28, [Online]. vol. 6, no. 1, pp. Available: https://journal.nurulfikri.ac.id/index.php/JIT
- M. Rifqi Zamzami, M. Riswandha İmawan, I. Ghozali, and U. [9] Muhammadiyah Surabaya, "A Comparative Study On Hadoop Ecosystem: Hive And HBase-A Literature Review," *Universitas* Pembangunan Nasional Veteran Jawa Timur, vol. 1, p. 3, 2024, doi: 10.3342/isstek.v2i1.26.
- N. Putu, A. Sintiawati, I. Bagus, and P. Purbadharmaja, "Analisis [10] Pengaruh Tenaga Kerja, Modal, Harga, dan Teknologi Terhadap Pendapatan UMKM di Kota Denpasar," 2024, [Online]. Available: https://ojs.unud.ac.id/index.php/EEB/index
- M. S. Hartawan, S. Rosyida, A. Hamid, W. Dari, and A. S. Putra, [11] Big Data (Informasi Dan Kasus). Jawa Timur: Tim Kun Fayakun,
- A. Lusiana and P. Yuliarty, "Penerapan Metode Peramalan (Forecasting) pada Permintaan Atap di PT X," 2020. [12]
- T. S. Nababan, "Efficiency and Elasticity of Labor Use on [13]
- Economic Sectors in Indonesia," 2019. G. Salimova, A. Ableeva, T. Lubova, Z. Zalilova, and A. [14] Sharafutdinov, "The Role of Agriculture in Gross Added Value," Montenegrin Journal of Economics, vol. 16, no. 1, pp. 183-191, 2020, doi: 10.14254/1800-5845/2020.16-1.12.
- [15] S. Cao, Y. Zeng, S. Yang, and S. Cao, "Research on Python Data Visualization Technology," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1757/1/012122.
- S. Sukparungsee, Y. Areepong, and R. Taboran, "Exponentially [16] weighted moving average—Moving average charts for monitoring the process mean," PLoS One, vol. 15, no. 2, Feb. 2020, doi: 10.1371/journal.pone.0228208.
- [17] I. Svetunkov and F. Petropoulos, "Old dog, new tricks: a modelling view of simple moving averages.
- [18] M. Riziq sirfatullah Alfarizi, M. Zidan Al-farish, M. Taufiqurrahman, G. Ardiansah, and M. Elgar, "PENGGUNAAN PYTHON SEBAGAI BAHASA PEMROGRAMAN UNTUK MACHINE LEARNING DAN DEEP LEARNING," 2023.
- A. A. Pratama, D. Agushinta R., and M. A. Mukhyi, "Penerapan [19] Metode Moving Average dan Exponential Smoothing untuk Prediksi Nilai Ekspor dan Impor Indonesia," Jurnal Ilmiah FIFO, vol. 14, no. 1, p. 58, May 2022, doi: 10.22441/fifo.2022.v14i1.006.
- Z. Kan, X. Cheng, S. H. Kim, and Y. Jin, "Apache Hive-Based Big [20] Data Analysis of HealthCare Data," 2018. [Online]. Available: http://www.acadpubl.eu/hub/
- [21] M. Marwan, Z. Saharuna, and R. Nur, "Performansi Response Time Query Pada Hadoop-Hive Menggunakan Metode Partition," Jurnal Teknologi Elekterika, vol. 5, no. 1, p. 19, May 2021, doi: 10.31963/elekterika.v5i1.2313.
- [22] N. Subagya, A. Wijajarto, and A. Almaarif, "Implementasi dan Analisis Hadoop Element Availability Berdasarkan Daemon Log Monitoring," 2021.
- N. Buslim, B. Busman, N. S. Sinatrya, and T. S. Kania, "Analisa [23] Sentimen Menggunakan Data Twitter, Flume, Hive Pada Hadoop dan Java Untuk Deteksi Kemacetan di Jakarta," Jurnal Online Informatika, vol. 3, no. 1, p. 1, Jun. 2018, doi: 10.15575/join.v3i1.141.
- L. Febtriani, "Studi Dan Perbandingan Apache Spark Sql dan Hive [24] Dalam Konteks Analisis Big Data," 2018.
- "AWS: What is Apache Hive?" [25]

- [26] Masbullah, "Implementasi Knowledge Management Untuk Mendorong Inovasi dan Produktivitas UMKM Menggunakan Big Data Media Sosial," *Jurnal Manajemen Informatika dan Bisnis Digital*, vol. 2, no. 1, pp. 12–23, 2023, doi: 10.55123/jumintal.v2i1.1887.
- [27] T. Bustomi, E. Marlovia, and Achdiat, "Inisiatif Open Data Pada Usaha Mikro Kecil Menengah Kota Bandung," *Jurnal Ilmu Administrasi*, vol. 12, 2021.
- [28] M. R. Fauzi, R. A. Pratama, P. Laksono, and P. Eosina, "Penerapan Big Data Menggunakan Algoritma Multi-Label K-Nearest Neighbor dalam Analisis Sentimen Konsumen UMKM Sektor Kuliner," *Krea-TIF*, vol. 9, no. 1, p. 9, May 2021, doi: 10.32832/kreatif.v9i1.3587.
- [29] Septa and Hoirul, "Peran Big Data pada Sektor Industri Perdagangan: Tinjauan Literatur pada Perusahaan Bidang Perkantoran Septa Hoirul," 2022, [Online]. Available: https://ejournal.unesa.ac.id/index.php/joa
- [30] E. Nazari, M. H. Shahriari, and H. Tabesh, "Big Data Analysis In Healthcare: Apache Hadoop, Apache Spark And Apache Flink," Frontiers in Health Informatics, vol. 8, 2019, doi: 10.30699/fbi.v8i1.180.
- [31] O. Azeroual and R. Fabre, "Processing big data with apache hadoop in the current challenging era of COVID-19," *Big Data and Cognitive Computing*, vol. 5, no. 1, 2021, doi: 10.3390/bdcc5010012.
- [32] S. H. Bhatti, A. Ahmed, A. Ferraris, W. M. Hirwani Wan Hussain, and S. F. Wamba, "Big data analytics capabilities and MSME innovation and performance: A double mediation model of digital platform and network capabilities," *Ann Oper Res*, 2022, doi: 10.1007/s10479-022-05002-w.
- [33] Kementrian Koordinator Bidang Perekonomian Republik Indonesia, "Perkembangan UMKM sebagai Critical Engine Perekonomian Nasional Terus Mendapatkan Dukungan Pemerintah,"

- https://www.ekon.go.id/publikasi/detail/4593/perkembangan-umkmsebagai-critical-engine-perekonomian-nasional-terusmendapatkan-dukungan-pemerintah, Jakarta, Oct. 01, 2022.
- [34] Cirprandy Riopaldo Tambunan, "Kontribusi UMKM dalam Perekonomian Indonesia," https://www.fajarharapan.id/2023/06/kontribusi-umkm-dalam-perekonomian-indonesia/, Pasaman, Jun. 26, 2023.
- N. Putri and J. T. Informasi, "Optimalisasi Terhadap Infrastruktur Big Data Dengan Menggunakan Hadoop," 2022.
- [36] C. Wibawa, S. Wirawan, M. Mustikasari, and D. T. Anggraeni, "Komparasi Kecepatan Hadoop Mapreduce Dan Apache Spark Dalam Mengolah Data Teks," *Jurnal Ilmiah MATRIK*, vol. 24, no. 1, 2022
- [37] I. R. Prabaswara and R. Saputra, "Analisis Data Sosial Media Twitter Menggunakan Hadoop dan Spark," IT JOURNAL RESEARCH AND DEVELOPMENT, vol. 4, no. 2, Mar. 2020, doi: 10.25299/itjrd.2020.vol4(2).4099.
- [38] I. D. N. A. Manuaba, I. B. G. Manuaba, and M. Sudarma, "Komparasi Metode Peramalan Grey dan Grey-Markov untuk mengetahui Peramalan PNBP di Universitas Udayana," *Majalah Ilmiah Teknologi Elektro*, vol. 21, no. 1, p. 83, Jul. 2022, doi: 10.24843/mite.2022.v21i01.p12.
- [39] Y. Pramana, R. S. Hartati, and K. Oka Saputra, "Peramalan Penerbitan Ijin Mendirikan Bangunan Dengan Single Moving Average Dan Exponential Smoothing," *Majalah Ilmiah Teknologi Elektro*, vol. 18, no. 2, Jul. 2019, doi: 10.24843/mite.2019.v18i02.p13.
- [40] I. G. N. G. A. Suniantara, N. Gunantara, and M. Sudarma, "Analisis Penyebaran Covid 19 Menggunakan Model SIR (Susceptible, Infected, Recovered) Di Provinsi Bali," *Majalah Ilmiah Teknologi Elektro*, vol. 22, no. 1, p. 39, Jun. 2023, doi: 10.24843/mite.2023.v22i01.p05.

Putu Denanta Bayuguna Perteka: Analisis Big Data UMKM...

9 772503 237160

{Halaman ini sengaja dikosongkan}