

DAFTAR ISI

| | |
|---|-------|
| <i>ABSTRACT</i> | i |
| ABSTRAK | ii |
| KATA PENGANTAR | iii |
| DAFTAR ISI..... | v |
| DAFTAR GAMBAR | viii |
| DAFTAR TABEL..... | xi |
| BAB I PENDAHULUAN | I-1 |
| 1.1. Latar Belakang | I-1 |
| 1.2. Rumusan Masalah | I-3 |
| 1.3. Batasan Masalah..... | I-4 |
| 1.4. Tujuan Penelitian..... | I-4 |
| 1.5. Manfaat Penelitian..... | I-4 |
| 1.6. Metodologi Penelitian | I-5 |
| 1.7. Sistematika Penulisan..... | I-6 |
| BAB II LANDASAN TEORI | II-1 |
| 2.1 Big Data..... | II-1 |
| 2.1.1. Data, Informasi, dan Pengetahuan | II-1 |
| 2.1.2. Big Data | II-2 |
| 2.1.3. Dimensi-Dimensi <i>Big Data</i> | II-3 |
| 2.1.4. Arsitektur <i>Big Data</i> | II-7 |
| 2.2. <i>Machine Learning</i> | II-11 |
| 2.3. <i>Support Vector Machine</i> | II-12 |
| 2.4. <i>Apache Spark</i> | II-13 |

| | | |
|-----------------------------------|---|-------|
| 2.5. | <i>Hadoop Mapreduce</i> | II-15 |
| 2.6. | Penelitian Terkait | II-15 |
| BAB III METODOLOGI..... | | III-1 |
| 3.1. | Studi Literatur..... | III-2 |
| 3.2. | Pengumpulan Kebutuhan | III-2 |
| 3.2.1. | Pengumpulan Dataset..... | III-2 |
| 3.2.2. | Pengumpulan Kebutuhan Perangkat | III-2 |
| 3.3. | Implementasi | III-4 |
| 3.3.1. | Proses Inisialisasi Data..... | III-4 |
| 3.3.2. | Visualisasi dan Analisis Data..... | III-5 |
| 3.3.3. | Penerapan Algoritma Support Vector Machine (SVM)..... | III-5 |
| 3.4. | Pengujian | III-5 |
| BAB IV HASIL DAN PEMBAHASAN | | IV-1 |
| 4.1 | Dataset | IV-1 |
| 4.2. | Perancangan Virtual Mesin | IV-3 |
| 4.2.1. | Konfigurasi Virtual Machine | IV-4 |
| 4.2.2. | Konfigurasi Apache Hadoop..... | IV-6 |
| 4.2.3. | Konfigurasi Apache Spark | IV-15 |
| 4.3. | Implementasi | IV-18 |
| 4.3.1. | Proses Inisialisasi Data..... | IV-18 |
| 4.3.2. | Visualisasi dan Analisis Data..... | IV-20 |
| 4.3.3. | Implementasi SVM dengan Grid Search..... | IV-45 |
| 4.4. | Pengujian | IV-47 |
| 4.4.1. | Pengujian SVM dengan Grid Search | IV-48 |
| 4.4.2. | Pengujian SVM dengan Perhitungan Manual | IV-76 |

| | |
|---------------------------------------|-------|
| 4.5. Pembahasan Hasil Penelitian..... | IV-99 |
| BAB V KESIMPULAN DAN SARAN..... | V-1 |
| 5.1. Kesimpulan..... | V-1 |
| 5.2. Saran..... | V-2 |
| DAFTAR PUSTAKA | |

DAFTAR GAMBAR

| | |
|---|-------|
| Gambar 2.1 Dimensi <i>Big Data</i> (Kannadasan et al., 2015)..... | II-3 |
| Gambar 2.2 <i>Traditional Information Architecture Capabilities</i> (Heller & Piziak, 2015) | II-8 |
| Gambar 2.3 <i>Big Data Information Architecture Capabilities</i> (Heller & Piziak, 2015) | II-8 |
| Gambar 2.4 <i>Oracle Integrated Information Architecture Capabilities</i> (Heller & Piziak, 2015) | II-9 |
| Gambar 2.5 <i>Machine Learning</i> (Pantech, 2018)..... | II-11 |
| Gambar 3.1 Diagram Alur Penelitian..... | III-1 |
| Gambar 4.1 Arsitektur Perancangan Sistem | IV-3 |
| Gambar 4.2 Konfigurasi OS Linux | IV-4 |
| Gambar 4.3 Menginstal tools yang digunakan..... | IV-5 |
| Gambar 4.4 Membuat keygen | IV-5 |
| Gambar 4.5 Menyalin keygen | IV-6 |
| Gambar 4.6 Menginstal Java..... | IV-6 |
| Gambar 4.7 Mengunduh file Apache Hadoop | IV-6 |
| Gambar 4.8 Memberikan hostname | IV-7 |
| Gambar 4.9 Meng-clone <i>Virtual Machine</i> | IV-7 |
| Gambar 4.10 Master host | IV-8 |
| Gambar 4.11 Slave 1 host | IV-8 |
| Gambar 4.12 Slave 2 host | IV-9 |
| Gambar 4. 13 Menyalin ssh yang sudah dibuat | IV-9 |

| | |
|--|-------|
| Gambar 4.14 Konfigurasi file core-site.xml | IV-9 |
| Gambar 4.15 Konfigurasi hdfs-site.xml..... | IV-11 |
| Gambar 4.16 Mendefinisikan worker | IV-11 |
| Gambar 4.17 Menyalin semua konfigurasi | IV-11 |
| Gambar 4.18 Melihat apakah HDFS sudah bekerja atau tidak | IV-12 |
| Gambar 4.19 Tampilan Hadoop pada browser | IV-12 |
| Gambar 4.20 Tampilan Data Node Slave 1 | IV-13 |
| Gambar 4.21 Tampilan Data Node Slave 2 | IV-13 |
| Gambar 4.22 Tampilan Yarn pada browser setelah dijalankan | IV-14 |
| Gambar 4.23 Mengunduh file Apache Spark..... | IV-15 |
| Gambar 4.24 Memasukkan environment Apache Spark..... | IV-16 |
| Gambar 4.25 Konfigurasi file Apache Spark..... | IV-16 |
| Gambar 4.26 Menambahkan worker node | IV-16 |
| Gambar 4.27 Tampilan Apache Spark pada browser..... | IV-17 |
| Gambar 4.28 Distribusi <i>smartphone</i> berdasarkan kapasitas baterai | IV-21 |
| Gambar 4.29 Distribusi <i>smartphone</i> berdasarkan <i>bluetooth</i> | IV-22 |
| Gambar 4.30 Distribusi <i>smartphone</i> berdasarkan kecepatan CPU | IV-23 |
| Gambar 4.31 Distribusi <i>smartphone</i> berdasarkan dual SIM..... | IV-24 |
| Gambar 4.32 Distribusi <i>smartphone</i> berdasarkan kamera depan..... | IV-25 |
| Gambar 4.33 Distribusi <i>smartphone</i> berdasarkan 4G..... | IV-26 |
| Gambar 4.34 Distribusi <i>smartphone</i> berdasarkan ketebalan..... | IV-27 |
| Gambar 4.35 Distribusi <i>smartphone</i> berdasarkan berat..... | IV-29 |
| Gambar 4.36 Distribusi <i>smartphone</i> berdasarkan jumlah inti | IV-30 |

| | |
|--|-------|
| Gambar 4.37 Distribusi <i>smartphone</i> berdasarkan kamera utama | IV-32 |
| Gambar 4.38 Distribusi <i>smartphone</i> berdasarkan tinggi resolusi | IV-33 |
| Gambar 4.39 Distribusi <i>smartphone</i> berdasarkan lebar layer | IV-35 |
| Gambar 4.40 Distribusi <i>smartphone</i> berdasarkan <i>RAM</i> | IV-36 |
| Gambar 4.41 Distribusi <i>smartphone</i> berdasarkan tinggi layar | IV-38 |
| Gambar 4.42 Distribusi <i>smartphone</i> berdasarkan lebar layar | IV-39 |
| Gambar 4.43 Distribusi <i>smartphone</i> berdasarkan ketahanan baterai | IV-41 |
| Gambar 4.44 Distribusi <i>smartphone</i> berdasarkan 3G | IV-42 |
| Gambar 4.45 Distribusi <i>smartphone</i> berdasarkan layar sentuh | IV-43 |
| Gambar 4.46 Distribusi <i>smartphone</i> berdasarkan <i>Wi-Fi</i> | IV-44 |
| Gambar 4.47 Kode untuk memplot <i>confusion matrix</i> | IV-45 |
| Gambar 4.48 Kode untuk proses prediksi dengan <i>SVM</i> | IV-46 |
| Gambar 4.49 Ilustrasi tabel <i>confusion matrix</i> | IV-77 |

DAFTAR TABEL

| | |
|---|-------|
| Tabel 2.1 Penelitian Terkait | II-17 |
| Tabel 2.2 Penelitian terdekat..... | II-22 |
| Tabel 3.1 Perangkat keras pengolah data..... | III-3 |
| Tabel 3.2 Perangkat lunak pengolah data | III-3 |
| Tabel 4. 1 Tabel parameter uji setiap pengujian | IV-47 |
| Tabel 4.2 Hasil plot confusion matrix pengujian 1 | IV-48 |
| Tabel 4.3 Nilai akurasi, testing time, dan training time pada pengujian ke-1 | IV-49 |
| Tabel 4.4 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-1 | IV-49 |
| Tabel 4.5 Hasil plot confusion matrix pengujian 2..... | IV-50 |
| Tabel 4.6 Nilai akurasi, testing time, dan training time pada pengujian ke-2 | IV-50 |
| Tabel 4.7 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-2 | IV-50 |
| Tabel 4.8 Hasil plot confusion matrix pengujian 3..... | IV-51 |
| Tabel 4.9 Nilai akurasi, testing time, dan training time pada pengujian ke-3 | IV-52 |
| Tabel 4.10 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-3 | IV-52 |
| Tabel 4.11 Hasil plot confusion matrix pengujian 4..... | IV-53 |

| | |
|---|-------|
| Tabel 4.12 Nilai akurasi, testing time, dan training time pada pengujian ke-4 | IV-53 |
| Tabel 4.13 Nilai recall, precision, F1-score, dan Support pada pengujian ke-4 | IV-54 |
| Tabel 4.14 Hasil plot <i>confusion matrix</i> pengujian 5 | IV-54 |
| Tabel 4.15 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-5 | IV-55 |
| Tabel 4.16 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-5 | IV-55 |
| Tabel 4.17 Hasil plot <i>confusion matrix</i> pengujian 6 | IV-56 |
| Tabel 4.18 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-6 | IV-56 |
| Tabel 4.19 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-6 | IV-57 |
| Tabel 4.20 Hasil plot <i>confusion matrix</i> pengujian 7 | IV-58 |
| Tabel 4.21 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-7 | IV-58 |
| Tabel 4.22 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-7 | IV-59 |
| Tabel 4.23 Hasil plot <i>confusion matrix</i> pengujian 8 | IV-59 |
| Tabel 4.24 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-8 | IV-60 |

| | |
|--|-------|
| Tabel 4. 25 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-8 | IV-60 |
| Tabel 4.26 Hasil plot <i>confusion matrix</i> pengujian 9 | IV-61 |
| Tabel 4.27 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-9..... | IV-61 |
| Tabel 4.28 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-9 | IV-62 |
| Tabel 4.29 Hasil plot <i>confusion matrix</i> pengujian 1 | IV-63 |
| Tabel 4.30 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-1 | IV-63 |
| Tabel 4.31 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-1 | IV-63 |
| Tabel 4.32 Hasil plot <i>confusion matrix</i> pengujian 2 | IV-64 |
| Tabel 4.33 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-2 | IV-65 |
| Tabel 4.34 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-2 | IV-65 |
| Tabel 4.35 Hasil plot <i>confusion matrix</i> pengujian 3 | IV-66 |
| Tabel 4.36 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-3 | IV-66 |
| Tabel 4.37 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-3 | IV-66 |
| Tabel 4.38 Hasil plot <i>confusion matrix</i> pengujian 4 | IV-67 |

| | |
|---|-------|
| Tabel 4.39 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-4 | IV-68 |
| Tabel 4.40 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-4 | IV-68 |
| Tabel 4.41 Hasil plot <i>confusion matrix</i> pengujian 5 | IV-69 |
| Tabel 4.42 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-5 | IV-69 |
| Tabel 4.43 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-5 | IV-69 |
| Tabel 4.44 Hasil plot <i>confusion matrix</i> pengujian 6 | IV-70 |
| Tabel 4.45 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-6 | IV-71 |
| Tabel 4.46 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-6 | IV-71 |
| Tabel 4.47 Hasil plot <i>confusion matrix</i> pengujian 7 | IV-72 |
| Tabel 4.48 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-7 | IV-72 |
| Tabel 4.49 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-7 | IV-73 |
| Tabel 4.50 Hasil plot <i>confusion matrix</i> pengujian 8 | IV-73 |
| Tabel 4.51 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-8 | IV-74 |

| | |
|---|-------|
| Tabel 4.52 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-8 | IV-74 |
| Tabel 4.53 Hasil plot <i>confusion matrix</i> pengujian 9 | IV-75 |
| Tabel 4.54 Nilai akurasi, <i>testing time</i> , dan <i>training time</i> pada pengujian ke-9 | IV-75 |
| Tabel 4.55 Nilai <i>recall</i> , <i>precision</i> , <i>F1-score</i> , dan <i>Support</i> pada pengujian ke-9 | IV-76 |
| Tabel 4.56 Hasil plot <i>confusion matrix</i> pengujian 1 | IV-78 |
| Tabel 4.57 Nilai TP, TN, FP, FN pengujian 1 | IV-78 |
| Tabel 4.58 Hasil plot <i>confusion matrix</i> pengujian 2 | IV-80 |
| Tabel 4.59 Nilai TP, TN, FP, FN pengujian 2 | IV-80 |
| Tabel 4.60 Hasil plot <i>confusion matrix</i> pengujian 3 | IV-82 |
| Tabel 4.61 Nilai TP, TN, FP, FN pengujian 3 | IV-83 |
| Tabel 4.62 Hasil plot <i>confusion matrix</i> pengujian 4 | IV-85 |
| Tabel 4.63 Nilai TP, TN, FP, FN pengujian 4 | IV-85 |
| Tabel 4.64 Hasil plot <i>confusion matrix</i> pengujian 5 | IV-87 |
| Tabel 4.65 Nilai TP, TN, FP, FN pengujian 5 | IV-87 |
| Tabel 4.66 Hasil plot <i>confusion matrix</i> pengujian 6 | IV-89 |
| Tabel 4.67 Nilai TP, TN, FP, FN pengujian 6 | IV-90 |
| Tabel 4.68 Hasil plot <i>confusion matrix</i> pengujian 7 | IV-92 |
| Tabel 4.69 Nilai TP, TN, FP, FN pengujian 7 | IV-92 |
| Tabel 4.70 Hasil plot <i>confusion matrix</i> pengujian 8 | IV-94 |
| Tabel 4.71 Nilai TP, TN, FP, FN pengujian 8 | IV-94 |

| | |
|--|--------|
| Tabel 4.72 Hasil plot confusion matrix pengujian 9 | IV-96 |
| Tabel 4.73 Nilai TP, TN, FP, FN pengujian 9 | IV-97 |
| Tabel 4.74 Hasil pengujian pada <i>Apache Spark</i> | IV-99 |
| Tabel 4.75 Hasil pengujian pada <i>Apache Hadoop</i> | IV-101 |