

DAFTAR ISI

HALAMAN PENGESAHAN	i
PRAKATA	v
DAFTAR ISI	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xiii
INTISARI	xiv
ABSTRACT	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Metodologi Penelitian	5
1.7 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	7
BAB III LANDASAN TEORI	12
3.1 BigData	12
3.2 Analisis BigData	13
3.3 Arsitektur Hadoop	13
3.3.1 Hadoop Distributed File System (HDFS)	14
3.3.2 MapReduce	14
3.3.3 YARN (Yet Another Resource Negotiator)	16
3.3.4 Spark	16
3.4 Data mining	18
3.4.1 Jenis- jenis Algoritme Data mining	19
3.4.2 Proses Data mining	20
3.5 Parquet	21
3.6 Min-Max Normalization	22
3.7 MLlib	22
3.8 Algoritme K-Means	23

3.9	Implementasi Paralel.....	24
3.9.1	Algoritme Scalable K-Means++	25
3.10	Euclidean Distance.....	26
3.11	Sum of Squared Error (SSE).....	26
3.12	Metode Elbow	27
3.13	Silhouette Coefficient	27
BAB IV ANALISIS DAN PERANCANGAN		29
4.1	Deskripsi Umum	29
4.2	Analisis Kebutuhan Sistem	29
4.3	Rancangan Penelitian.....	30
4.3.1	Business Understanding.....	30
4.3.2	Data Understanding	33
4.3.3	Data Preparation.....	36
4.3.4	Modeling	37
4.3.5	Evaluation	39
4.3.6	Deployment.....	39
BAB V IMPLEMENTASI		40
5.1	Lingkungan Implementasi	40
5.2	Implementasi Platform Spark	40
5.3	Implementasi Praproses Data.....	42
5.3.1	Praproses Data	42
5.3.2	Prapemrosesan Join.....	48
5.4	Implementasi Algoritme K-Means dan Scalable K-Means++	49
5.4.1	Algoritme K-Means	50
5.4.2	Algoritme Scalable K-Means++	50
5.5	Implementasi Pengujian SSE.....	51
5.5.1	Implementasi Penentuan Jumlah Klaster	51
5.6	Implementasi Pengujian Silhouette coefficient.....	54
5.7	Implementasi Analisis Data Hasil Clustering	54
5.8	Implementasi Visualisasi.....	56
5.8.1	Transformasi Data.....	56
5.8.2	Visualisasi Data.....	60
BAB VI HASIL DAN PEMBAHASAN		63

6.1	Hasil Pengujian Algoritme K-Means dan Scalable K-Means++	63
6.1.1	Perbandingan Algoritme Berdasarkan Nilai SSE.....	63
6.1.2	Perbandingan Algoritme berdasarkan Silhouette coefficient.....	66
6.2	Evaluasi Waktu Pemrosesan	68
6.3	Data Hasil Clustering.....	70
6.4	Hasil Visualisasi	72
BAB VII KESIMPULAN DAN SARAN		75
7.1	Kesimpulan	75
7.2	Saran	77
DAFTAR PUSTAKA.....		78
LAMPIRAN.....		81