



Pemanfaatan Teknik Clustering untuk Mencapai Kecerdasan Bisnis Menggunakan Algoritma K-Means

Rizki Dwi Romadhona¹

¹Magister Terapan Teknik Elektro, Computer Science Department, Politeknik Negeri Malang
email: rizki.dwi@polinema.ac.id¹

ABSTRAK

Kata Kunci:

K-means
Clustering
Kecerdasan Bisnis

Strategi bisnis dan kebijakan dalam suatu perusahaan atau organisasi pasti memiliki profit oriented dan mempunyai visi misi yang jelas terhadap pelayanan konsumennya. Penggunaan teknologi yang dapat membantu proses bisnis salah satunya adalah kecerdasan bisnis clustering yang dapat digunakan, membutuhkan teknik data yang tepat dan akurat untuk mencapai hasil yang maksimal menemukan solusi yang tepat. Permasalahan clustering adalah memperkirakan nilai atau jumlah k yang akan digunakan yang tepat untuk memperoleh hasil yang akurat dan mengestimasi nilai k terbaik. Penulisan ini memberikan pandangan terhadap efektifitas clustering menggunakan k-means untuk kegiatan bisnis yang dijalani seperti penjualan produk, periklanan, promosi penjualan. Setelah itu mengenali pola pola yang akan digunakan sebagai informasi dari sekumpulan koleksi data untuk menjadi bahan penarik kesimpulan. Oleh karena itu dalam penulisan ini ingin menunjukkan pemanfaatan teknik clustering untuk mencapai kecerdasan bisnis menggunakan algoritma k-means.

ABSTRACT

Keyword:

K-means
Clustering
Business Intelligence

Business strategies and policies in a company or organization must have a profit oriented and have a clear vision and mission of customer service. The use of technology that can help business processes, one of which is business intelligence clustering that can be used, requires appropriate and accurate data techniques to achieve maximum results in finding the right solution. The clustering problem is estimating the value or number of k that will be used appropriately to obtain accurate results and estimate the best k value. This paper provides an overview of the effectiveness of clustering using k-means for business activities undertaken such as product sales, advertising, sales promotion. After that, identify the patterns that will be used as information from a collection of data to become material for drawing conclusions. Therefore, in this paper, we want to show the use of clustering techniques to achieve business intelligence using the k-means algorithm.



PENDAHULUAN

Dalam dunia bisnis yang terjadi saat ini mempunyai tingkat dimanis yang lebih maju dan tinggi untuk mencapai tingkat profit yang lebih besar. Pelaku usaha harus menemukan berbagai cara - cara terbaik dan efisien untuk mengembangkan usaha dan bisnis mereka karena persaingan semakin ketat [1]. Perusahaan atau organisasi pasti menggunakan data yang telah dikumpulkan kemudian dikelompokkan untuk itu menggunakan data mining yang digunakan perusahaan sangat banyak dan memakan waktu yang lama tanpa adanya teknik yang baik untuk digunakan pengelompokan data tersebut sehingga adanya *clustering* dengan metode *k-means* ini dapat membantu mereka untuk peningkatan efisien dalam mengelompokkan data [1]. Data *mining* adalah suatu konsep yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. Data *mining* merupakan proses semi otomatis yang menggunakan teknik matematis, kecerdasan buatan dan *machine learning* untuk mengidentifikasi informasi bermanfaat yang ada dalam *database* [2]. Data *mining* menurut Connolly adalah proses menemukan informasi yang benar tanpa diketahui sebelumnya dan dapat dipahami serta bermanfaat dari basis data besar (*Big Data*) yang dapat mendukung keputusan penting [3]. Menurut penelitian [4] *clustering* merupakan teknik pengolahan data *mining* yang termasuk pada kategori *unsupervised learning* yaitu pembelajaran tanpa guru atau tanpa menggunakan label kelas seperti pada klarifikasi. *Clustering* pada algoritma *k-means* mencari nilai kemiripan antar data dengan menghitung jarak salah satunya menggunakan *euclidian distance*.

Kecerdasan bisnis menggunakan algoritma k-means sudah sangat banyak digunakan salah satunya adalah untuk mempermudah pelaku usaha menemukan solusi dengan melakukan proses pengumpulan dan analisa informasi yang ada untuk membuat keputusan yang terbaik dalam menjalankan bisnis [5]. Analisa informasi bagi pengusaha diperoleh dari retailer yang langsung berhubungan dengan konsumen. Dari data retail akan diperoleh pola belanja konsumen dan kecenderungan lainnya yang akan meningkatkan keuntungan dan efisiensi pengusaha [5]. Pada penentuan pola pasar, *Customer Relationship Management* (CRM) penelitian [6] memaparkan perlu adanya perhatian untuk menentukan teknik data mining pada dimensi *customer identification* teknik yang tepat adalah secara berurutan yaitu, *classification*, *clustering*, *regression* dimana *classification* dan *clustering* memiliki selisih atau perbedaan penelitian.

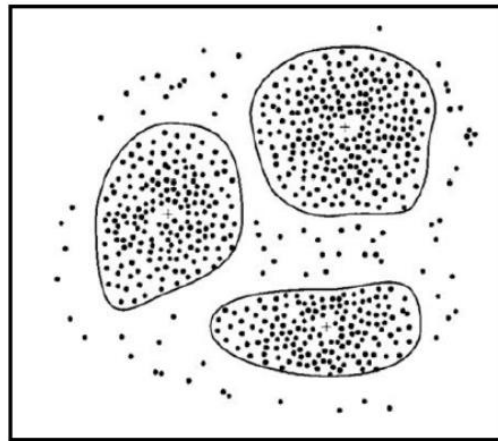
Perpaduan antara *clustering* dan kecerdasan bisnis sudah diterapkan salah satunya oleh [7] pada klusterisasi konsumen untuk menentukan *high-profit*, *high value*, dan *low-risk customer*. Hasil dari penulisan paper ini adalah untuk mengetahui seberapa efektif dari penggunaan pemanfaatan metode teknik clustering untuk mencapai kecerdasan bisnis mendapatkan hasil k terbaik untuk menentukan keputusan dan

strategi terbaik yang harus dilakukan sehingga mendapatkan keuntungan yang maksimal sebagai landasan pelaku usaha atau para pengusaha.

METODE

Metode *clustering* dalam data *mining* termasuk metode yang sudah cukup dikenal dan banyak dipakai dalam data *mining*. Hingga saat ini para peneliti masih melakukan perbaikan model *clustering* karena metode yang dikembangkan sekarang masih bersifat *heuristic* [8]. Usaha yang sudah dilakukan sejauh ini untuk menghitung jumlah *cluster* yang maksimal dan paling baik masih dilakukan namun demikian menggunakan metode yang sekarang tidak menjamin hasil pengklasteran yang optimal. Tapi biasanya sudah memenuhi dan sudah cukup bagus untuk pencapaian dari segi praktis.

Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data atau objek ke dalam grup sehingga setiap cluster akan berisi data yang homogen atau sejenis dengan data yang masuk di dalamnya seperti gambar 2.1 [9]. Dalam metode *clustering* berusaha untuk mendekatkan objek yang mirip dengan jarak yang paling terdekat dalam satu grup atau *cluster* dan membuat jarak antar *cluster* sejauh mungkin, yang berarti objek dalam satu cluster memiliki satu kemiripan antar satu sama lain yang membuat mereka menjadi satu grup dan berbeda dengan cluster – cluster lain [10]. Dalam pengaplikasian metode ini tidak diketahui sebelumnya untuk menentukan jumlah cluster dan seperti apa pengelompokannya.



Gambar 2.1 Ilustrasi *Clustering* [9]

Clustering menggunakan algoritma *K-means* merupakan suatu algoritma yang mudah digunakan di dalam teknik pengelompokan karena mempunyai prakiraan yang efisien dan tidak memerlukan parameter yang banyak. Algoritma *K-means* menggunakan k kelompok yang digunakan sebagai



centroid dan secara beriterasi melalui proses pengiraan titik tengah sehingga suatu fungsi kriteria tersebut telah dicapai dan tetap [11]. Di dalam teknik pengelompokan, perkiraan untuk membedakan satu kelompok dengan yang lain menggunakan satu algoritma yang dipanggil fungsi jarak yaitu setiap tahap persamaan atau perbedaan [12]. Menurut [8] Berikut langkah – langkah untuk menuju *clustering* menggunakan algoritma k-means:

Pilih jumlah *cluster* k .

Inisialisasi k pusat *cluster*, pusat-pusat *cluster* diberi nilai awal dengan angka random.

Alokasikan semua data/objek ke *cluster* terdekat. Kedekatan ditentukan jarak demikian juga dengan data ke *cluster* tertentu ditentukan dengan jarak antar data dengan pusat *cluster*. Kemudian menghitung jarak dengan Euclidean yang dirumuskan:

$$D(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2 \dots}$$

Dimana:

$D(i,j)$ = jarak antara i ke pusat cluster j

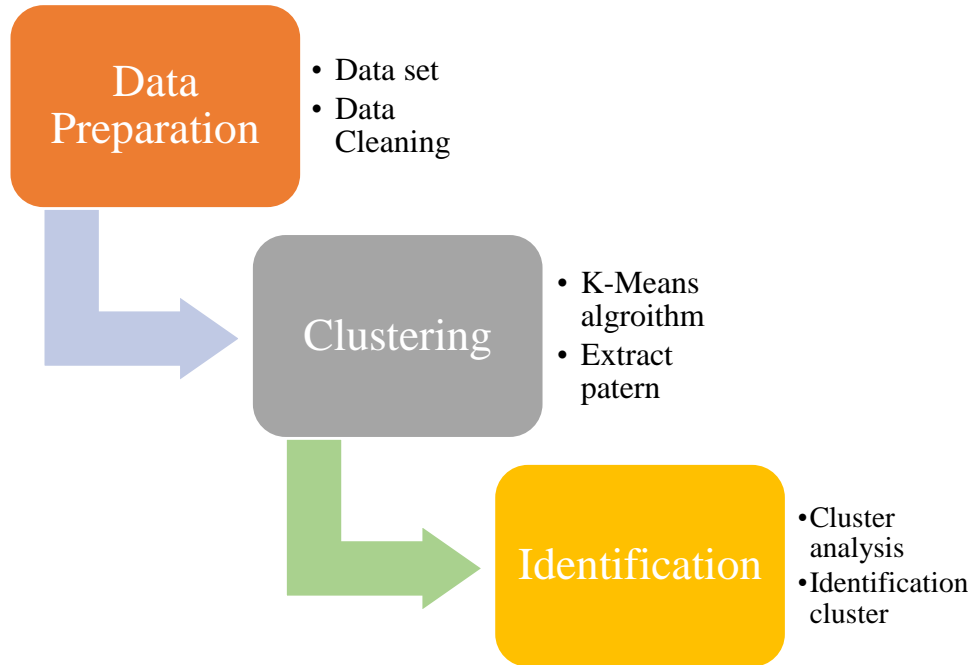
X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang.

Kerjakan lagi setiap objek memakai pusat cluster baru. Jika *cluster* tidak berubah lagi maka proses *clustering* selesai.

Berdasarkan penelitian yang telah dilakukan oleh [7] maka metode penelitian paper ini dituliskan dengan beberapa teknik untuk menentukan nilai k dalam k-means yang akan digunakan dalam clustering yang memiliki 3 tahapan yaitu: data preparation, clustering dan identification.



Gambar 2.2 Proses *Clustering* [13]

Data preparation yaitu pemilihan data set dan data *cleaning*. Data set yang dipilih adalah yang terkait dengan kecerdasan bisnis. Data *cleaning* adalah penghapusan data yang tidak penting atau tidak termasuk yang dibutuhkan [13]. Sedangkan *clustering* ada tahapan yaitu menggunakan algoritma *k-means* untuk menentukan jarak antar data yang akan dijadikan satu grup sesuai dengan kemiripan data dan karakteristik data tersebut [13]. Kemudian dijadikan identifikasi untuk menemukan berapa jumlah *cluster* dan tergolong menjadi berapa grup yang terbentuk dari pola pola tersebut.

HASIL DAN PEMBAHASAN

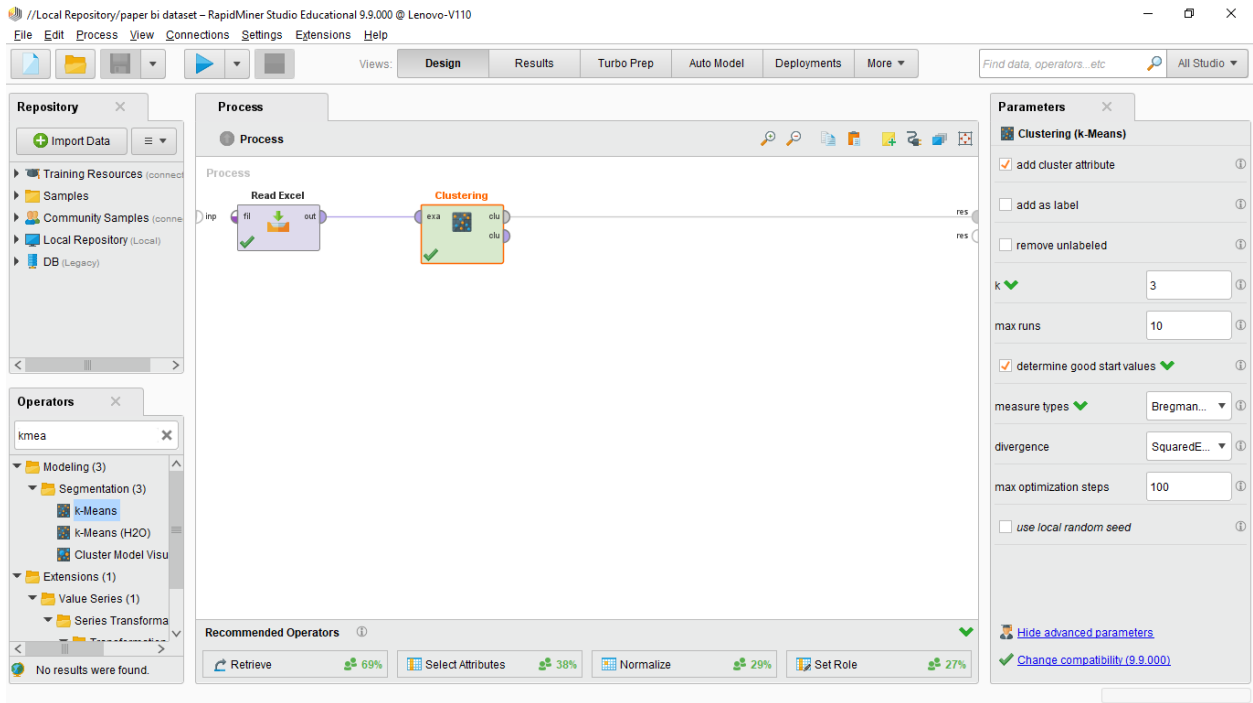
Dalam pembahasan penelitian ini menggunakan data set yang merupakan data *preparation* tahap ini melakukan bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan atau dijalankan. Pertama tahapan yang dikerjakan adalah melakukan perbaikan pengabaian atribut pada data mentah yang dianggap tidak relevan atau tidak kongruen dengan hasil pengujian dan perubahan terhadap nilai data pada atribut dataset dengan tujuan untuk mempermudah pemahaman terhadap isi *record*. Data set yang digunakan adalah data contoh dari produksi tebu beberapa daerah di Jawa Timur seperti di bawah ini:



Tabel 3.1 Data Produksi Tebu

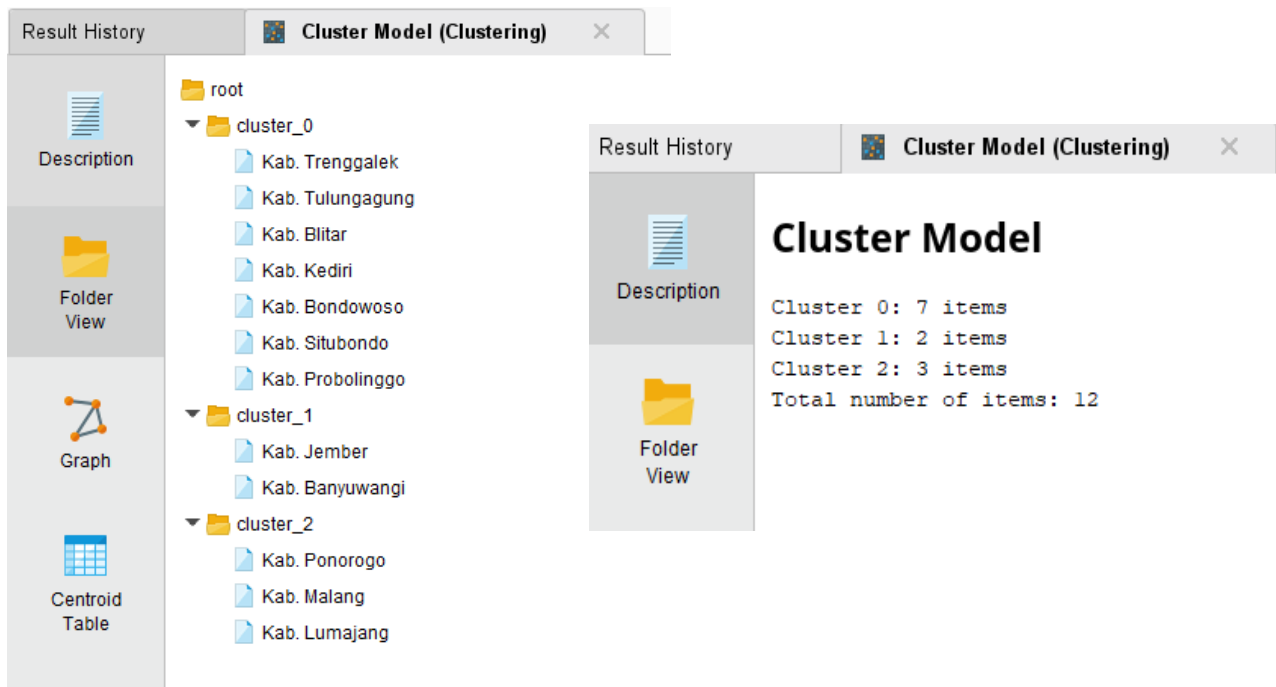
NO	Kota /Kab	Luas Lahan	Produksi
1	Kab. Ponorogo	66693	402047
2	Kab. Trenggalek	31136	182848
3	Kab. Tulungagung	49230	259581
4	Kab. Blitar	50577	289494
5	Kab. Kediri	51083	281392
6	Kab. Malang	65597	464498
7	Kab. Lumajang	72552	387168
8	Kab. Jember	162619	964001
9	Kab. Banyuwangi	113609	706419
10	Kab. Bondowoso	61330	329557
11	Kab. Situbondo	48902	290954
12	Kab. Probolinggo	59130	311258

Dari data tersebut diketahui jumlah *Cluster* ada 3, jumlah data ada 12 dan jumlah atribut ada 2 (Luas Lahan dan Produksi). Setelah itu data diatas akan dicari jumlah produksi yang tertinggi dan terendah untuk dijadikan pengelompokannya menggunakan algoritma *k-means* metode jarak Euclidean dibantu dengan software *rapid miner*. Gambar 3.1 menunjukkan design rapid miner dimana read excel diisi dengan data produksi dan *clustering* diseting menggunakan nilai $k = 3$ dan *max runs* 10 serta *divergence* menggunakan *Euclidean distance*.



Gambar 3.1 Permodelan *K-Means* pada *Rapid Miner*

Setelah dilakukan run program maka muncul hasil *cluster* atau pengelompokan grafik, *centroid table*, *folder view* seperti berikut:



Gambar 3.2 Hasil Cluster Berdasarkan Jumlah Produksi Daerah



Attribute	cluster_0	cluster_1	cluster_2
Luas Lahan	50198.286	138114	68280.667
Produksi	277869.143	835210	417904.333

Gambar 3.3 hasil *clustering* menggunakan *k-means* nilai $k=3$

Kedua gambar diatas menunjukan angka dan penggolongan daerah yang terbagi menjadi 3 cluster berdasarkan jarak terdekat dari nilai produksi dengan luas lahan yang menjadi atributnya. *Cluster 1* terdapat meliputi daerah yang memiliki tingkat produksi yang rerataan yang tinggi yaitu Kab. Jember dan Kab. Banyuwangi sedangkan pada *cluster 2* yang mempunyai rerataan tertinggi kedua yaitu Kab. Ponorogo, Kab. Malang, dan Kab. Lumajang untuk cluster 0 adalah hasil produksi yang paling bawah dengan luasan wilayah yang tertera pada data set excel di atas.

SIMPULAN

Dari hasil penelitian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

Setelah dilakukan pengelompokan data produksi daerah sesuai dengan luas lahan produksi menggunakan algoritma *K-means clustering* yang diolah dengan *rapid miner* terbentuk 3 cluster yaitu: Cluster 1 adalah cluster jumlah rerataan produksi tertinggi yang terjadi pada daerah Kab. Jember dan Kab. Banyuwangi dengan nilai rerataan sebesar 835210

Cluster 2 adalah cluster jumlah rerataan produksi tertinggi kedua yang terjadi pada Kab. Ponorogo, Kab. Malang, dan Kab. Lumajang dengan nilai rerataan sebesar 417904,33

Cluster 0 adalah jumlah rerataan produksi terendah dengan nilai sebesar 277869,143

Strategi produksi tebu dari daerah yang terindeks paling rendah bisa ditingkatkan dari hasil cluster yang telah dilakukan dengan menjadikan clustering ini sebagai acuan tingkat produksi selanjutnya.

DAFTAR RUJUKAN

- [1] Sutrisno, Afriyudi, Widiyanto, PENERAPAN DATA MINING PADA PENJUALAN MENGGUNAKAN METODE CLUSTERING STUDY KASUS PT. INDOMARCO PALEMBANG. Universitas Bina Dharma Palembang, Palembang, November 2011.



- [2] Dewi Arianti Wulandari, Hendra Jatnika, Yudhy S. Purwanto, RANCANG BANGUN APLIKASI CLUSTERING DATA MINING MENGGUNAKAN METODE K-MEANS DAN K-MODES. Teknik Informatika, Sekolah Tinggi Teknik PLN Jakarta, April 2018.
- [3] Connolly (2002, p1115). Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data. USA: Pearson Education Limited.
- [4] Folorunso O, Ogunde A O, et.al. Data Mining for Business Intelligence in Distribution Chain Analytics. International Journal of Computer, the Internet and Management. 2010; 18(1): 15-26.
- [5] Gupta G and Agrawal H. Improving Customer Relationship Management using Data Mining. International Journal of Machine Learning and Computing. 2012; 2(6): 874-877
- [6] Rodpysh K V, Aghai A. Applying Data Mining in Customer Relationship Management. International Journal of Information Technology Control and Automation (IJITCA). 2012; 2(3): 15-25.
- [7] Rajagopal. Sankar. (2011). Customer Data Clustering Using Data Mining Technique, International Journal of Database Management Systems, Vol.3, No.4.
- [8] Santosa, B. (2007). Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisni Yogyakarta: Graha Ilmu.
- [9] Suprihatin, (2011). Klastering K-Means Untuk Penentuan Nilai Ujian, JUSI, Vol. 1, No. 1.
- [10] Larose, Daniel T, (2006). Data Mining Methods and Models. Hoboken New Jersey: Jhon Wiley & Sons, Inc,
- [11] Ong. Johan Oscar. (2013). Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University, Jurnal Ilmiah Teknik Industri, Vol. 12, No. 1
- [12] Arifin.Muhammad. (2014). “Bussiness Intelligence Untuk Prediksi Customer Churn Telekomunikasi”, Prosiding SNATIF, Vol.1, Agusta, Y., K-means - Penerapan, Permasalahan dan Metode Terkait. Jurnal Sistem dan Informatika Vol. 3: 47-60.
- [13] Fajrizal, Khairani Djahara, “Pemanfaatan Clustering Algoritma McKmeans untuk Business Intelligence”, Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Lancang Kuning, Konferensi Nasional Sistem & Informatika 2015.