

PENENTUAN JUMLAH CLUSTER IDEAL SMK DI JAWA TENGAH DENGAN METODE X-MEANS CLUSTERING DAN K-MEANS CLUSTERING

Rifki Adhitama¹, Auliya Burhanuddin², Ridho Ananda³

¹Program Studi Rekayasa perangkat Lunak, Institut Teknologi Telkom Purwokerto

²Program Studi Informatika, Institut Teknologi Telkom Purwokerto

³Program Teknik Industri, Institut Teknologi Telkom Purwokerto

Email: ¹rifki@ittelkom-pwt.ac.id, ²auliya@ittelkom-pwt.ac.id, ³ridho.ananda@ittelkom-pwt.ac.id

(Naskah masuk: 19 Februari 2020, diterima untuk diterbitkan: 13 Maret 2020)

Abstrak

SMK merupakan salah satu instrumen penting dalam pengembangan Sumber Daya Manusia (SDM) di Indonesia pada umumnya dan di Jawa Tengah pada khususnya. Belum adanya pengelompokan SMK berdasarkan data pokok kemendikbud di Jawa Tengah merupakan sebuah peluang untuk mengembangkan arah revitalisasi SMK menjadi lebih baik dan jelas. X-means merupakan salah satu metode clustering yang dikembangkan dari metode clustering yang cukup populer, yaitu K-means. Penelitian ini menggunakan data pokok kemendikbud untuk menghitung pembagian cluster terbaik dengan menggunakan metode X-means dengan membandingkan nilai Davis Buldin Index (DBI) X-means dengan nilai DBI K-means pada variasi ukuran cluster mulai dari empat, enam, delapan dan sepuluh cluster. Hasil penelitian ini menunjukkan bahwa secara konsisten nilai DBI terbaik ada pada ukuran cluster empat, baik menggunakan X-means maupun K-means dengan nilai DBI X-means sebesar 0,933 dan nilai DBI K-means sebesar 0,914, sedangkan nilai DBI paling besar juga konsisten pada ukuran cluster 10, sebesar 1,439 pada X-means dan 1,322 pada K-means. Berdasarkan hasil tersebut maka SMK di Jawa Tengah dapat dibagi ke dalam 4 kelompok yaitu kurang, cukup, baik, dan unggul.

Kata kunci: SMK, Clustering, X-Means, K-Means, Davies Bouldin Index

DETERMINING VOCATIONAL IDEAL CLUSTER NUMBER IN CENTRAL JAVA WITH X-MEANS CLUSTERING AND K-MEANS CLUSTERING METHODS

Abstract

Vocational School is one of the important instruments in the development of Human Resources (HR) in Indonesia in general and in Central Java in particular. The absence of vocational grouping based on basic data from the Ministry of Education and Culture of Indonesia in Central Java is an opportunity to develop a better and clearer direction of revitalization of Vocational School. X-means is one of the clustering methods developed from a quite famous clustering method, namely K-means. This study uses the Ministry of Education and Culture's basic data to calculate the best cluster distribution using the X-means method by comparing the Davis Buldin Index (DBI) X-means value with K-means DBI values on variations in cluster sizes ranging from four, six, eight and ten clusters. The results of this study indicate that consistently the best DBI values exist in cluster four, both using X-means or K-means with X-means DBI values of 0.933 and K-means DBI values of 0.914, while the highest DBI values are also consistent with cluster 10 size, 1,439 for X-means and 1,322 for K-means. Based on these results, Vocational Schools in Central Java can be divided into 4 groups, namely less, sufficient, good, and superior.

Keywords: Vocational School, Clustering, X-Means, K-Means, Davies Bouldin Index.

1. PENDAHULUAN

Perkembangan Sumber Daya Manusia (SDM) merupakan salah satu faktor krusial dalam menghadapi persaingan secara global, khususnya di Indonesia. Salah satu upaya untuk memenuhi kebutuhan Sumber Daya Manusia dengan menumbuhkan SDM yang berkualitas melalui

Pendidikan kejuruan. Pendidikan kejuruan adalah bagian dari system Pendidikan yang mempersiapkan seseorang agar lebih mampu bekerja pada satu kelompok pekerjaan atau satu bidang pekerjaan daripada bidang pekerjaan yang lainnya [1]. Sedangkan menurut penjelasan Undang-Undang Nomor 20 Tahun 2003 Pasal 15, pendidikan kejuruan merupakan pendidikan menengah yang

mempersiapkan peserta didik terutama untuk bekerja dalam bidang tertentu. Salah satu sarana pengembangan pendidikan kejuruan di Indonesia adalah Sekolah Menengah Kejuruan (SMK) [2].

Pengembangan dan revitalisasi SMK di Indonesia merupakan hal yang penting, hal ini sejalan dengan Instruksi Presiden (Inpres) Nomor 9 Tahun 2016 tentang perlunya revitalisasi SMK agar dapat meningkatkan kompetensi, produktivitas dan daya saing bangsa Indonesia. Hal ini sejalan dengan tujuan dari pemerintah provinsi Jawa Tengah yang merupakan salah satu provinsi dengan kepadatan penduduk tertinggi di Jawa [3] terus berusaha mengembangkan Pendidikan kejuruan di Jawa Tengah.

Salah satu tantangan dalam mewujudkan pengembangan SMK yang merata di Jawa Tengah adalah belum tersedianya pengelompokan secara terukur tentang sebaran kualitas SMK dan masih sedikitnya SMKN di Jawa Tengah [4] berdasarkan kriteria dari data pokok kemendikbud yaitu peserta didik, rombongan, guru, pegawai, ruang kelas, ruang laboratorium dan ruang perpustakaan.

Mengelompokkan persebaran kualitas SMK di Jawa Tengah berdasarkan kriteria tersebut dapat dilakukan dengan menggunakan teknik clustering. Karena pada data pokok SMK se-Jawa Tengah dari depdikbud. Salah satu teknik clustering yang sangat populer digunakan adalah k-means, salah satu pengembangan dari metode k-means yang dapat digunakan adalah x-Means clustering. Metode x-Means digunakan karena metode ini dapat diimplementasikan untuk dataset yang tidak terlalu besar [5], selain itu x-Means dapat membagi data ke dalam beberapa cluster secara optimal [6].

Penelitian ini bertujuan untuk menentukan berapa jumlah cluster ideal untuk membagi data SMK yang ada di Jawa Tengah berdasarkan data dari kemendikbud yang ada pada pangkalan data pokok kemendikbud. Data yang digunakan adalah data SMK di seluruh kabupaten dan kota di Jawa Tengah yang berjumlah 1458 SMK. Kriteria yang digunakan dalam penelitian ini adalah kriteria yang melekat pada data kemendikbud yaitu peserta didik, jumlah rombongan belajar, jumlah guru dan pegawai, jumlah ruang kelas, perpustakaan dan laboratorium.

Selanjutnya paper ini akan menjelaskan lebih detail tentang penelitian yang telah dilakukan sebagai berikut: bagian 2 menjelaskan tentang kerangka teori dimana bagian 3 menjelaskan tentang metode klastering yang digunakan dalam penelitian ini, bagian 4 hasil dan pembahasan lalu di bagian 5 adalah kesimpulan dari penelitian ini.

2. METODE PENELITIAN

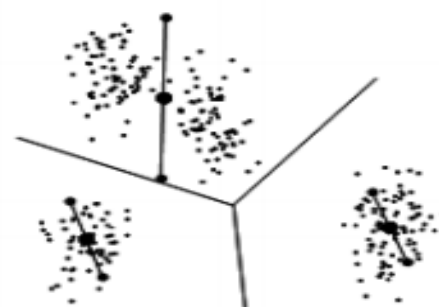
2.1. X-Means Clustering

Clustering secara umum didefinisikan sebagai pengelompokan objek-objek yang memiliki kemiripan berdasarkan kriteria tertentu [7]. Seperti

satu grup profile yang dibagi ke dalam beberapa cluster yang memiliki kemiripan satu sama lain. Clustering memiliki output cluster yang lebih sedikit atau sama dengan jumlah data profile masukan. [8]. Analisis klaster merupakan salah satu teknik pembelajaran mesin yang memiliki tujuan dasar untuk menemukan pola yang informatif atau pola yang berguna pada big data [9]. Metode k-Means merupakan salah satu metode yang populer digunakan untuk clustering data. Penelitian mengenai pembagian persebaran data dengan metode k-Means telah dilakukan sebelumnya untuk pengelompokan persebaran data kesehatan puskesmas di Indonesia [10], selain digunakan untuk pembagian sebaran data kesehatan metode k-means juga digunakan dalam pengelompokan hasil tangkapan ikan di pelabuhan ternate [11]. Penelitian penggunaan metode k-means juga digunakan untuk mengelompokkan jenis musik seperti musik klasik, rap, metal dan musik indian. Hasil dari penelitian tersebut adalah pengelompokan menggunakan metode k-means memiliki waktu komputasi yang cukup lama [12].

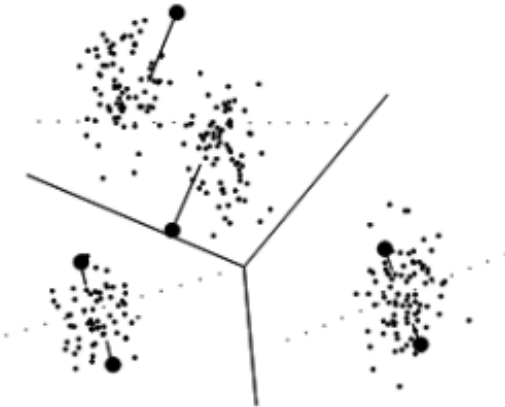
Metode k-means meskipun menghasilkan hasil yang cukup baik dan populer digunakan dalam pengelompokan data, metode ini masih memiliki beberapa kelemahan yaitu biaya komputasi yang tinggi, input nilai k secara manual [5]. Penelitian yang dilakukan oleh Pelleg pada tahun 2000 melakukan ekstensi kepada metode k-means untuk menghasilkan proses komputasi yang lebih cepat dan melakukan estimasi jumlah klaster yang paling efisien dalam pemrosesan data. Hasil dari penelitian tersebut adalah metode x-means (extended k-means) dapat digunakan pada dataset yang kecil dan proses komputasi yang lebih cepat dibandingkan dengan k-means [5].

Algoritma X-means terdiri dari beberapa operasi berulang hingga eksekusi berakhir, dasar proses dari algoritma x-means adalah memodifikasi struktur cluster k-means dengan membagi data ke dalam minimal dua cluster dilanjutkan dengan pembagian dua kluster untuk setiap klaster yang telah terbentuk (dua cluster lokal untuk setiap cluster). Jika menggunakan 3 cluster maka masing-masing cluster akan dibagi ke dalam dua lokal cluster dari pusat cluster awalnya.



Gambar 1. Pembagian masing-masing cluster ke dalam dua buah cluster lokal

Dua hasil cluster lokal ini bergerak berlawanan arah dalam ukuran region searah dengan vektor yang dipilih secara acak.

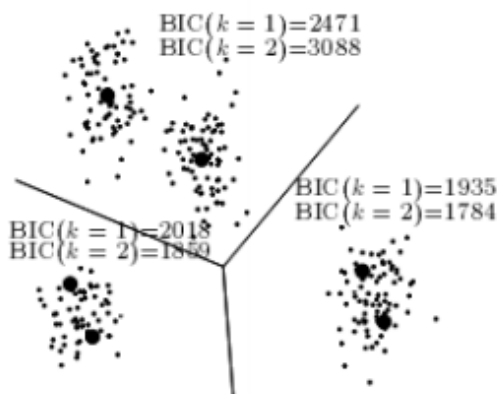


Gambar 2. Pembagian cluster lokal dengan arah vector acak searah dengan ukuran region

Setelah vektor dari cluster lokal berhasil dibentuk langkah selanjutnya adalah menghitung nilai Bayesian Information Criterion (BIC) dari cluster awal (yang belum dibagi menjadi dua kluster lokal) dengan nilai BIC setelah dibagi ke dalam dua cluster lokal perhitungan nilai BIC ditunjukkan dalam persamaan (1).

$$BIC(M_j) = \hat{l}_j(D) - \frac{P_j}{2} \cdot \log R \quad (1)$$

Dimana $\hat{l}_j(D)$ adalah fungsi log-likelihood data berdasarkan modek ke- j dan P_j merupakan jumlah parameter pada M_j , diketahui juga sebagai Schwarz Criterion. Jika nilai BIC setelah pembagian cluster lokal lebih tinggi daripada nilai BIC cluster awal, maka cluster tersebut akan dibagi menjadi dua cluster. Jika nilai BIC cluster awal lebih tinggi dari BIC setelah dibagi menjadi cluster lokal maka cluster tersebut akan tetap nilai (tidak dibagi menjadi dua cluster).



Gambar 3. Perbandingan nilai BIC antara cluster awal dengan cluster setelah dibagi ke dalam dua cluster lokal

Iterasi akan terus berulang sampai nilai BIC cluster sebelumnya lebih baik daripada nilai cluster setelah dibagi ke dalam dua cluster lokal.



Gambar 4. Hasil akhir jumlah cluster setelah dibandingkan nilai BIC-nya

2.2. Davies Buldin index

Davies bouldin index (DBI) merupakan salah satu metode yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin. Davies Bouldin Index digunakan untuk mengevaluasi cluster secara umum [13] berdasarkan kuantitas dan kedekatan antar anggota cluster [14]. Perhitungan nilai Davies Bouldin Index berdasarkan perbandingan rasio cluster ke- i dan cluster ke- j . Semakin kecil nilai Davies Bouldin Index maka semakin baik cluster yang dihasilkan [14]. Perhitungan nilai DBI disajikan pada persamaan (2).

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (2)$$

2.3 Dataset

Pada penelitian ini bahan yang digunakan adalah data SMK se-Jawa tengah yang didapatkan dengan cara pengambilan data dari Data Pokok Pendidikan Dasar dan Menengah Direktorat Jenderal Pendidikan Dasar dan Menengah Kementerian Pendidikan dan Kebudayaan pada daring <http://dapo.dikdasmen.kemdikbud.go.id>. Data yang dikumpulkan adalah berupa data SMK di Jawa Tengah beserta parameter-parameter kunci yang dapat digunakan untuk analisis pada penelitian ini. Beberapa parameter tersebut antara lain peserta didik, rombongan belajar, jumlah guru, pegawai, ruang kelas, ruang laboratorium dan ruang perpustakaan.

2.4 Prosedur Penelitian

Prosedur yang dilakukan dalam penelitian ini meliputi beberapa tahapan antara lain:

a. *Identifikasi masalah*

Analisis masalah pada penelitian ini dilakukan berdasarkan berdasarkan situs data pokok kemendikbud, dimana belum adanya pengelompokan kualitas SMK yang ada khususnya di Jawa Tengah, meskipun pada situs data pokok Kemendikbud tersebut telah ada kriteria kriteria yang ditetapkan untuk masing-masing jenjang Pendidikan khususnya SMK.

b. *Pengambilan data*

Setelah didapatkan hasil dari identifikasi masalah, maka penelitian dilanjutkan pada pengumpulan bahan penelitian yaitu data yang diambil dari data pokok Kemendikbud terkait dengan data SMK di Jawa Tengah. Karena pada pangkalan data kemendikbud data tidak terkelompok per provinsi, pengambilan data dilakukan dengan mensortir setiap SMK yang ada di setiap kabupaten /Kota yang ada di Jawa Tengah.

c. *Analisis data awal dan pembersihan data*

Setelah bahan penelitian didapatkan tahapan penelitian selanjutnya adalah analisis data awal dan pembersihan data. Pada tahapan ini peneliti melakukan analisis awal dari data yang telah didapatkan, misal untuk kriteria apakah dapat digunakan semua dan datanya telah ternormalisasi atau belum. Selain itu juga dilakukan pembersihan data terkait apakah terdapat data yang memiliki kriteria kosong atau tidak sesuai sebelum dilanjutkan pada tahapan selanjutnya. Data yang digunakan hanya data semester Ganjil 2018/2019 yang memiliki jumlah sinkronisasi data lebih dari 10 kali. Contoh dataset dan kriteria yang digunakan dalam penelitian ini dapat dilihat pada Gambar 5. Dimana

PD = Peserta didik
Rombel = Rombongan Belajar
Guru = Total jumlah guru pada SMK yang bersangkutan
Pegawai = Jumlah pegawai pada SMK yang bersangkutan
R. Kelas = Ruang Kelas
R. Lab = Ruang Laboratorium
R. Perpustakaan = Ruang Perpustakaan

d. *Pemrosesan data*

Tahapan selanjutnya dalam penelitian ini adalah melakukan klastering dari data yang telah didapatkan pada tahapan sebelumnya dengan menggunakan metode X-Means Clustering.

e. *Analisis hasil dan evaluasi*

Tahapan analisis hasil merupakan tahapan terakhir pada penelitian ini dimana hasil x-means clustering ini nantinya akan di analisis hasil cluster dengan Davies Bouldin Index untuk menghasilkan sebuah rekomendasi dan hasil Analisa terhadap persebaran kualitas SMK di Jawa Tengah berdasarkan kriteria yang terdapat pada data pokok Kemendikbud.

| Jml Sync | PD | Rombel | Guru | Pegawai | R. Kelas | R. Lab | R. Perpustakaan |
|----------|-------|--------|------|---------|----------|--------|-----------------|
| 74 | 1.610 | 45 | 85 | 23 | 35 | 9 | 1 |
| 91 | 1.855 | 52 | 94 | 29 | 35 | 8 | 1 |
| 54 | 114 | 6 | 8 | 2 | 6 | 1 | 1 |

Gambar 5. Contoh data yang diambil dari pangkalan data Kemendikbud

2.5 Skenario pengujian

Skenario pengujian yang dilakukan dalam penelitian ini meliputi beberapa skenario. Skenario pertama yaitu membagi jumlah cluster ke dalam beberapa bagian, yaitu pembagian cluster awal pada x-means dengan ukuran 4, 6, 8, dan 10 cluster. Skenario kedua yaitu menggunakan metode k-means standar dan menghitung nilai Davies Bouldin Index (DBI) dari masing masing cluster bentukan yang telah dibagi pada skenario pertama dan skenario kedua.

3. HASIL DAN PEMBAHASAN

Tabel 1. Nilai Davis Bouldin Index untuk setiap ukuran cluster pada x-means

| Ukuran Cluster | Davis Bouldin Index (DBI) |
|----------------|---------------------------|
| 4 | 0,933 |
| 6 | 1,211 |
| 8 | 1,291 |
| 10 | 1,439 |

Tabel 1 menunjukkan hasil dari Davies Bouldin Index untuk setiap ukuran cluster. Berdasarkan tabel tersebut, nilai Davies Bouldin Index berbanding lurus dengan ukuran cluster. Davies Bouldin index paling rendah didapatkan pada ukuran cluster 4 sebesar 0,933 dan paling tinggi didapatkan pada ukuran cluster 10 sebesar 1,439.

Nilai DBI paling baik yang dihasilkan pada skenario penelitian pertama sebesar 0,933 pada ukuran cluster 4. Sedangkan untuk skenario penelitian kedua DBI terbaik dihasilkan pada ukuran cluster 4 dengan nilai DBI 0,914. Hasil skenario penelitian kedua ditunjukkan pada Tabel 2.

Tabel 2. Nilai Davis Bouldin Index untuk setiap ukuran cluster pada k-means

| Ukuran Cluster | Davis Bouldin Index (DBI) |
|----------------|---------------------------|
| 4 | 0,914 |
| 6 | 1,125 |
| 8 | 1,293 |
| 10 | 1,322 |

Hasil nilai DBI pada kedua skenario penelitian menunjukkan hasil yang konsisten yaitu nilai DBI paling baik ada pada ukuran cluster 4. Hal ini disebabkan persebaran data SMK yang digunakan merupakan data terstruktur yang rentang persebaran datanya tidak terlalu luas dengan kriteria yang tidak terlalu banyak.

4. KESIMPULAN

Hasil dari kedua skenario penelitian yang telah dilakukan menunjukkan nilai DBI yang konsisten antara X-means dan K-means clustering untuk mengukur hasil cluster persebaran data pokok SMK dengan enam buah kriteria. Hasil DBI terbaik ada pada ukuran cluster 4 baik pada skenario penelitian pertama maupun skenario penelitian kedua. Hasil DBI paling buruk ada pada ukuran cluster 10 baik pada skenario penelitian pertama maupun skenario penelitian kedua.

Hasil yang konsisten tersebut dihasilkan data yang digunakan merupakan data terstruktur dan rentang data yang digunakan tidak terlalu jauh untuk setiap kriterianya. Sehingga sebaran SMK yang ada di Jawa Tengah berdasarkan data pokok kemendikbud dapat dibagi ke dalam 4 kelompok yang dapat dibagi ke dalam kategori kurang, cukup, baik dan unggul.

Berdasarkan pengelompokan tersebut nantinya diharapkan dapat memberikan gambaran kepada *stakeholder* ataupun pemerintah Jawa Tengah agar dapat memetakan ataupun menyusun instrumen yang cocok digunakan untuk pengembangan dan revitalisasi SMK di Jawa Tengah berdasarkan jumlah ideal dari kriteria SMK yang dihasilkan dari penelitian ini.

Pengukuran nilai cluster dengan metode lain seperti dunn index dapat digunakan pada penelitian selanjutnya untuk menilai hasil cluster dari algoritma yang digunakan pada penelitian ini. Selai. Itu penelitian selanjutnya dapat menggunakan data tidak terstruktur atau data yang rentang data pada setiap kriterianya lebar. Sehingga didapatkan pakah rentang data, dan jumlah kriteria berpengaruh signifikan terhadap hasil bentukan cluster berdasarkan nilai DBI.

5. DAFTAR PUSTAKA

- [1] R. Evans. 1987. "Fondation of Vocational Education." Olympus Publishing Company, Salt Lake City.
- [2] T. Damarjati. 2016. "Artikel: Konsep Pembelajaran di Sekolah menengah Kejuruan." *Direktorat Pembinaan SMK Direktorat Jenderal Pendidikan Dasar dan Menengah Kementerian Pendidikan dan Kebudayaan*. [Online]. Available: <http://psmk.kemdikbud.go.id/konten/1869/konsep-pembelajaran-di-sekolah-menengah-kejuruan>.
- [3] Badan Pusat Statistik. 2010. "Halaman Utama," *Badan Pusat Statistik: Sensus Penduduk 2010*. [Online]. Available: <https://sp2010.bps.go.id/>.
- [4] Bidang IKP Humas Jateng. 2018. "Archive: Berita Publik: Jateng Terus Revitalisasi Pendidikan Vokasi untuk Hadapi Revolusi Industri 4.0.," *Humas Pemprov jaten*. [Online]. Available: <https://jatengprov.go.id/publik/jateng-terus-revitalisasi-pendidikan-vokasi-untuk-hadapi-revolusi-industri-4-0/>.
- [5] D. Pelleg and A. Moore. 2002. "X-Means: Extending K-Means with Efficient Estimation if the Number of Clusters." Pittsburgh.
- [6] A. G. Jondya and B. H. Iswanto. 2017. "Indonesian's Traditional Music Clustering Based on Audio Features," *Procedia Comput. Sci.*, vol. 116, pp. 174–181.
- [7] R. O. Duda and P. E. Hart. 1973. *Pattern classification and scene analysis*, 1st editio. New York: Wiley.
- [8] A. Al-wakeel and J. Wu. 2016. "K-means based cluster analysis of residential smart meter measurements," *Energy Procedia*, vol. 88, pp. 754–760.
- [9] I. H. Witten and E. Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques (Google eBook)*, Second Edi. Morgan Kaufmann Publishers.
- [10] A. Burhanuddin, F. M. Wibowo, and A. Arif Prasetyo. 2016. "Pengelompokkan Sebaran Tenaga Kesehatan Puskesmas Di Indonesia Dengan Menggunakan Pengklasteran K-Means," in *Konferensi Nasional Sistem Informasi*, pp. 1257–1263.
- [11] R. Hablum, A. Khairan, and R. Rosihan. 2019. "Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means," *JIKO (Jurnal Inform. dan Komputer)*, vol. 2, no. 1, pp. 26–33.
- [12] A. Sen, 2014. "Automatic Music Clustering using Audio Attributes," *Int. J. Comput. Sci. Eng.*, vol. 3, no. 06, pp. 307–312.
- [13] S. Nawrin, M. Rahatur, and S. Akhter. 2017. "Exploreing K-Means with Internal Validity Indexes for Data Clustering in Traffic Management System," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 3.
- [14] A. Bates and J. Kalita. 2016. "Counting clusters in twitter posts," in *ACM International Conference Proceeding Series*, vol. 04-05-Marc.