

PENERAPAN ALGORITMA NAIVE BAYES DAN PARTICLE SWARM OPTIMIZATION UNTUK KLASIFIKASI BERITA HOAX PADA MEDIA SOSIAL

Risa Wati

Program Studi Sistem Informasi Akuntansi
Universitas Bina Sarana Informatika
www.bsi.ac.id
risawati.rwx@bsi.ac.id

Abstract— *Social media is the most effective way to facilitate fast information, unfortunately there are some elements who use social media to add hoax or deception to give misleading opinions to the public. Therefore a method is needed to classify hoax news and non hoax news on social media. Naive Bayes is a simple classification algorithm but has high qualifications, but Naive Bayes has a very sensitive shortcomings in the selection of features and therefore the Particle Swarm Optimization method is needed to improve the expected results. After conducting research with the Naive Bayes method and the Naive Bayes method based on Particle Swarm Optimization, the results obtained are Naive Bayes yielding 74.67% while the Naive Bayes based on Particle Swarm Optimization with an accuracy value of 85.19%. The purpose of this study is to see a large comparison. Swarm Optimization particles to improve accuracy in the classification of hoax news on social media using the Naive Bayes classifier. After using Particle Swarm Optimization the test results increased by 10.52%.*

Keywords: *Classification, Social Media, Naive Bayes.*

Intisari— *Media sosial adalah cara paling efektif untuk menyebarkan informasi secara cepat, sayang sekali ada beberapa oknum yang memanfaatkan media sosial untuk menginformasikan berita bohong atau hoax untuk memberikan opini yang menyesatkan kepada masyarakat. Maka dari itu dibutuhkan metode untuk mengklasifikasikan berita hoax dan berita non hoax pada media sosial. Naive Bayes merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki akurasi yang tinggi, namun Naive Bayes memiliki kekurangan yaitu sangat sensitive dalam pemilihan fitur maka dari itu dibutuhkan metode Particle Swarm Optimization untuk meningkatkan hasil akurasi. Setelah dilakukan penelitian dengan metode Naive Bayes dan metode Naive Bayes berbasis Particle Swarm Optimization maka hasil yang didapat adalah Naive Bayes menghasilkan akurasi sebesar 74.67% sedangkan Naive Bayes berbasis Particle*

Swarm Optimization nilai akurasi sebesar 85.19%. Tujuan dari penelitian ini adalah untuk melihat seberapa besar pengaruh Particle Swarm Optimization untuk meningkatkan akurasi pada klasifikasi berita hoax pada media sosial menggunakan pengklasifikasi Naive Bayes. Setelah menggunakan Particle Swarm Optimization hasil akurasi meningkat sebesar 10.52%.

Kata Kunci: *Klasifikasi, Media Sosial, Naive Bayes.*

PENDAHULUAN

Media sosial adalah cara paling efektif untuk menyebarkan informasi secara cepat. Banyak orang menjadikan sosial media sebagai tempat untuk memperoleh informasi. Ada beberapa pengguna yang memanfaatkan media sosial untuk hal yang baik dan ada juga yang memanfaatkan media sosial untuk kepentingan yang tidak baik. Dampak negatif yang disebabkan oleh penggunaan media sosial salah satunya adalah sebagai tempat penyebaran kabar bohong atau hoax (Kussa & Utama, 2018). Penyebaran berita bohong atau hoax pada media sosial oleh sejumlah oknum yang tidak bertanggung jawab dapat memancing perselisihan dan lain sebagainya.

Pada saat ini kita sering kesulitan untuk membedakan mana berita yang benar dan mana berita bohong atau hoax pada media sosial (Nurchayati, 2019). Jika tidak berhati-hati maka kita akan mudah termakan tipuan berita hoax dan bahkan ikut serta menyebarkan berita hoax tersebut (Detiknews, 2017). Maka dari itu dibutuhkan algoritma untuk menentukan berita tersebut benar atau hoax. Klasifikasi berita hoax atau berita dengan informasi yang tidak benar merupakan salah satu aplikasi kategorisasi teks (Rasywir & Purwarianti, 2015)

Beberapa penelitian sebelumnya yang telah dilakukan adalah klasifikasi berita hoax pilpres menggunakan metode K-NN dan pembobotan menggunakan TF-IDF (Rozi & Sulistyawati, 2019),

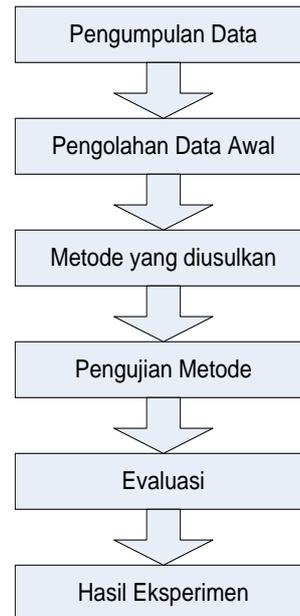


metode klasifikasi Rocchio untuk analisis hoax (Afriza & Adisantoso, 2019), klasifikasi konten berita dengan metode text mining (Kurniawan, Effendi, & Sitompul, 2012) dan masih banyak lagi penelitian sebelumnya yang telah dilakukan. Metode Naive Bayes dan K-Nearest Neighbor adalah dua metode yang sering digunakan dalam klasifikasi teks (Anggono, Suryani, & Kurniati, 2009). Naive Bayes merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki akurasi yang tinggi, namun Naive Bayes memiliki kekurangan yaitu sangat sensitive dalam pemilihan fitur sehingga dapat mempengaruhi hasil akurasi. Metode klasifikasi berbasis fitur yang dikembangkan dalam penelitian-penelitian tersebut menghasilkan akurasi yang baik (Kussa & Utama, 2018). Particle Swarm Optimization (PSO) adalah algoritma populer dan bionik berdasarkan perilaku sosial yang terkait dengan burung berkelompok untuk masalah optimasi (Yan & Jiao, 2016).

Dengan demikian pada penelitian ini pengklasifikasi Naive Bayes dengan optimasi Particle Swarm Optimization akan diterapkan untuk klasifikasi berita hoax pada media sosial.

BAHAN DAN METODE

Sumber data dari penelitian ini adalah data kualitatif. Dataset yang digunakan adalah berita dari media sosial yang diperoleh dari <https://turnbackhoax.id/> dimana berita tersebut terdiri dari 75 berita Hoax dan 75 berita non Hoax. Dataset yang akan diolah berbentuk sekumpulan teks yang terpisah dalam bentuk dokumen. Data berita hoax dan non hoax dikumpulkan dan disimpan dengan nama hoax dan non_hoax dalam media teks yaitu notepad dengan ekstensi .txt. Penelitian ini menggunakan metode eksperimen, Metode yang dipilih pada penelitian ini berdasarkan kesesuaian data dengan metode pengklasifikasi teks yang paling baik yang sudah digunakan oleh beberapa peneliti sebelumnya. Metode yang digunakan adalah algoritma Naive Bayes dan untuk meningkatkan optimasi yaitu menggunakan Particle Swarm Optimization. Untuk eksperimen data penelitian, peneliti menggunakan RapidMiner 5.3 untuk mengolah data dan alat bantu mengukur akurasi dari eksperimen. Langkah-langkah pada penelitian dapat dilihat pada gambar dibawah ini:



Sumber: (Wati, 2019)

Gambar 1. Langkah Penelitian

HASIL DAN PEMBAHASAN

Setelah data dikumpulkan dan disimpan dalam notepad dengan ekstensi .txt. langkah selanjutnya dalam penelitian ini adalah melakukan preprocessing atau pengolahan awal data melalui tiga tahapan proses yaitu sebagai berikut:

1. Tokenization

Yaitu mengumpulkan semua kata dan menghilangkan tanda baca maupun simbol yang bukan huruf, seperti “ , . / : ;) dan lain-lain. Contoh hasil pengolahan tokenization pada RapidMiner dapat dilihat pada tabel dibawah ini:

Tabel 1. Perbandingan Teks Sebelum Dan Sesudah Dilakukan Proses Tokenization

Teks Sebelum Dilakukan Proses Tokenization	Teks Setelah Dilakukan Proses Tokenization
Pihak Pemerintah Provinsi Nusa Tenggara Barat (NTB) dan Duta Besar RI di Korea Selatan memberikan klarifikasi terkait kabar terlanturnya mahasiswa NTB di Korea Selatan. Gubernur NTB Dr. Zulkieflimansyah membantah kabar mahasiswa penerima beasiswa ke Korea Selatan terlanter. Ia menyatakan kondisi mahasiswa di Korea saat ini masih dalam tahap kursus bahasa dan belum aktif berkuliah.	Pihak Pemerintah Provinsi Nusa Tenggara Barat NTB dan Duta Besar RI di Korea Selatan memberikan klarifikasi terkait kabar terlanturnya mahasiswa NTB di Korea Selatan Gubernur NTB Dr Zulkieflimansyah membantah kabar mahasiswa penerima beasiswa ke Korea Selatan terlanter Ia menyatakan kondisi mahasiswa di Korea saat ini masih dalam tahap kursus bahasa dan belum aktif berkuliah

Sumber: (Wati, 2019)

2. Stopwords Removal

Stopwords dapat diartikan sebagai menghilangkan kata-kata umum yang tidak memiliki makna atau informasi yang dibutuhkan. Contoh hasil pengolahan stopwords removal pada RapidMiner dapat dilihat pada tabel dibawah ini:

Tabel 2. Perbandingan Teks Sebelum Dan Sesudah Dilakukan Proses Stopwords Removal

Teks Sebelum Dilakukan Proses Stopwords Removal	Teks Setelah Dilakukan Proses Stopwords Removal
Pihak Pemerintah Provinsi Nusa Tenggara Barat (NTB) dan Duta Besar RI di Korea Selatan memberikan klarifikasi terkait kabar terlanturnya mahasiswa NTB di Korea Selatan. Gubernur NTB Dr. Zulkieflimansyah membantah kabar mahasiswa penerima beasiswa ke Korea Selatan terlanter. Ia menyatakan kondisi mahasiswa di Korea saat ini masih dalam tahap kursus bahasa dan belum aktif berkuliah.	Pihak Pemerintah Provinsi Nusa Tenggara Barat NTB dan Duta Besar RI di Korea Selatan memberikan klarifikasi terkait kabar terlanturnya mahasiswa NTB di Korea Selatan Gubernur NTB Dr Zulkieflimansyah membantah kabar mahasiswa penerima beasiswa ke Korea Selatan terlanter Ia menyatakan kondisi mahasiswa di Korea saat ini masih dalam tahap kursus bahasa dan belum aktif berkuliah

Sumber: (Wati, 2019)

3. Stemming

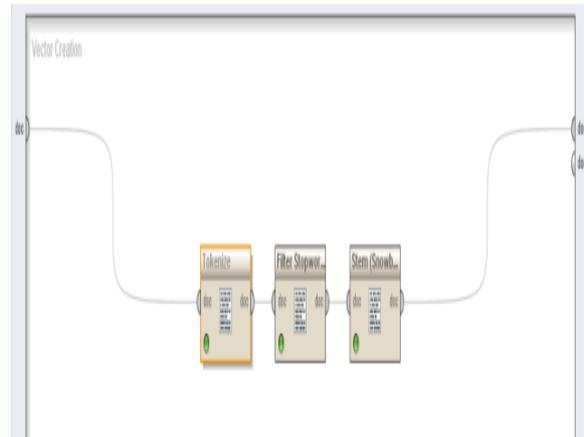
Stemming merupakan salah satu proses dari mengubah token yang berimbuhan menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada token tersebut. Pentingnya stemming dalam proses pembuatan sistem adalah untuk menghilangkan imbuhan pada awalan dan akhiran. Contoh hasil pengolahan stemming pada RapidMiner dapat dilihat pada tabel dibawah ini:

Tabel 3. Perbandingan Teks Sebelum Dan Sesudah Dilakukan Proses Stemming

Teks Sebelum Dilakukan Proses Stemming	Teks Setelah Dilakukan Proses Stemming
Pihak Pemerintah Provinsi Nusa Tenggara Barat (NTB) dan Duta Besar RI di Korea Selatan memberikan klarifikasi terkait kabar terlanturnya mahasiswa NTB di Korea Selatan. Gubernur NTB Dr. Zulkieflimansyah membantah kabar mahasiswa penerima beasiswa ke Korea Selatan terlanter. Ia menyatakan kondisi mahasiswa di Korea saat ini masih dalam tahap kursus bahasa dan belum aktif berkuliah.	pihak pemerintah provinsi nusa tenggara barat ntb dan duta besar ri di korea selatan memberikan klarifikasi terkait kabar terlanturnya mahasiswa ntb di korea selatan gubernur ntb dr zulkieflimansyah membantah kabar mahasiswa penerima beasiswa ke korea selatan terlanter ia menyatakan kondisi mahasiswa di korea saat ini masih dalam tahap kursus bahasa dan belum aktif berkuliah

Sumber: (Wati, 2019)

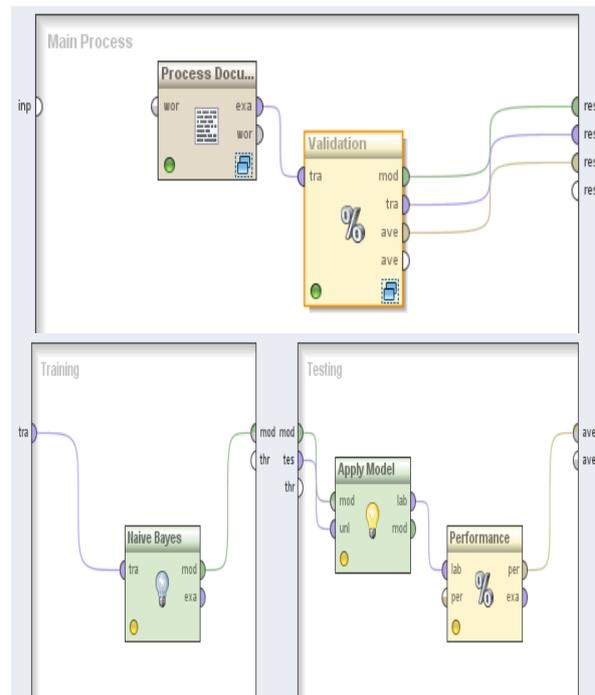
Preprocessing pada RapidMiner dapat dilihat pada gambar berikut ini:



Sumber: (Wati, 2019)

Gambar 2. Preprocessing

Data training yang digunakan terdiri dari 75 data berita hoax dan 75 data berita non hoax pada media sosial. Pada penelitian ini hasil klasifikasi berita hoax dan non hoax pada media sosial menggunakan metode Naive Bayes pada framework RapidMiner sebagai berikut



Sumber: (Wati, 2019)

Gambar 3. Model Pengujian Validasi Naive Bayes

Untuk data berita non hoax, 57 diklasifikasikan kedalam berita non hoax dan 18 data diprediksi berita non hoax ternyata hasil prediksinya merupakan berita hoax. Untuk data berita hoax, 55 diklasifikasikan sebagai berita hoax



dan 20 data diprediksi kedalam berita hoax ternyata hasil prediksianya merupakan berita non hoax. Hasil klasifikasi menggunakan Naive Bayes dapat dilihat pada gambar dibawah ini:

accuracy: 74.67% +/- 12.83% (mikro: 74.67%)			
	true non_hoax	true hoax	class precision
pred non_hoax	57	20	74.03%
pred hoax	18	55	75.34%
class recall	76.00%	73.33%	

Sumber: (Wati, 2019)

Gambar 4. Model Confusion Matrix untuk Metode Naive Baye

$$1. Accuracy = \frac{TN+TH}{TN+FN+TP+FP} \dots\dots\dots(1)$$

$$Accuracy = 0.7467 (74,67\%)$$

$$2. Sensitivity = \frac{TH}{TH+FN} \dots\dots\dots(2)$$

$$Sensitivity = 0.7534 (75.34\%)$$

$$3. Specificity = \frac{TN}{TN+FH} \dots\dots\dots(3)$$

$$Specificity = 0.7403 (74.03\%)$$

$$4. PPV = \frac{TH}{TH+FH} \dots\dots\dots(4)$$

$$PPV = 0.7333 (73.33\%)$$

$$5. NPV = \frac{TN}{TN+FN} \dots\dots\dots(5)$$

$$NPV = 0.76 (76\%)$$

Pengujian Naive Bayes berbasis Particle Swarm Optimization, nilai training dalam penelitian ini ditentukan dengan cara melakukan uji coba dengan memasukkan nilai parameter Population Size dan Inertia Weight. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk menentukan nilai training.

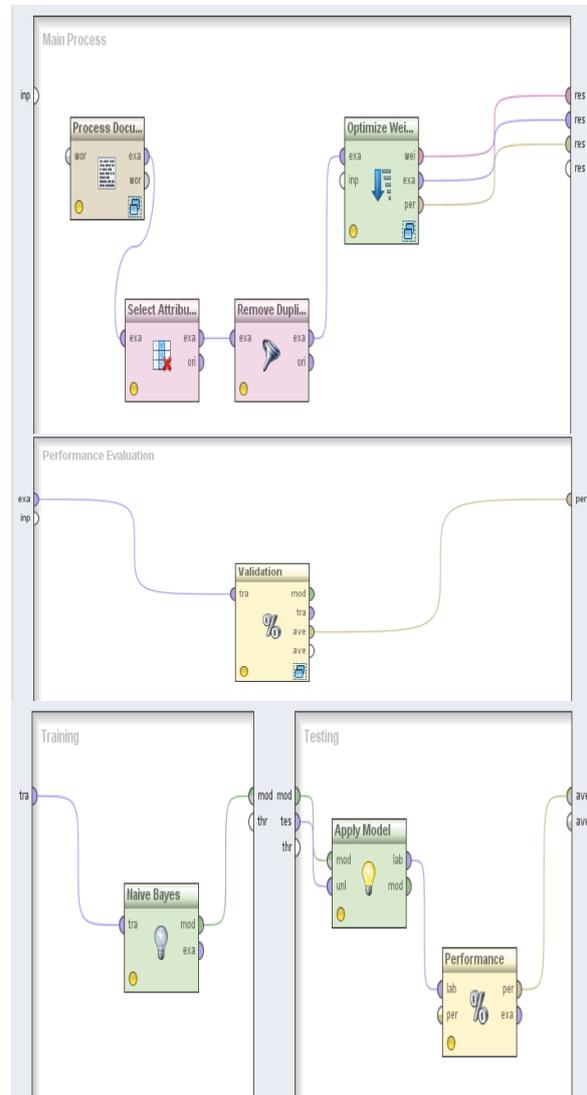
Tabel 4. Eksperimen Penentuan Nilai Training NB berbasis PSO

Population size (Q)	Inertia Weight (w)	Accuracy
5	1.0	81.10%
6	1.0	85.19%
7	1.0	81.71%
6	0.1	83.14%

Sumber: (Wati, 2019)

Hasil terbaik pada eksperimen Naive Bayes berbasis Particle Swarm Optimization adalah dengan nilai Population Size sebesar 6 dan nilai

Inertia Weight sebesar 1.0, Accuracy yang dihasilkan adalah sebesar 85.19%. Pada penelitian ini hasil klasifikasi berita hoax dan non hoax pada media sosial menggunakan metode Naive Bayes berbasis Particle Swarm Optimization pada framework RapidMiner sebagai berikut



Sumber: (Wati, 2019)

Gambar 5. Model Pengujian Validasi Naive Bayes Berbasis Particle Swarm Optimization

Untuk data berita non hoax, 63 diklasifikasikan kedalam berita non hoax dan 10 data diprediksi berita non hoax ternyata hasil prediksianya merupakan berita hoax. Untuk data berita hoax, 57 diklasifikasikan sebagai berita hoax dan 18 data diprediksi kedalam berita hoax ternyata hasil prediksianya merupakan berita non hoax. Hasil Pengujian Model Naive Bayes berbasis PSO dapat dilihat pada gambar dibawah ini



accuracy: 85.19% +/- 0.27% (mikro: 85.14%)

	true nonhoax	true hoax	class precision
pred. nonhoax	82	11	84.93%
pred. hoax	11	64	85.33%
class recall	84.93%	85.33%	

Sumber: (Wati, 2019)

Gambar 6. Model Confusion Matrix untuk Metode Naive Bayes Berbasis PSO

$$1. Accuracy = \frac{TN+TH}{TN+FN+TH+FH} \dots\dots\dots (6)$$

$$Accuracy = 0.8514 (85.14\%)$$

$$2. Sensitivity = \frac{TH}{TH+FN} \dots\dots\dots (7)$$

$$Sensitivity = 0.8533 (85.33\%)$$

$$3. Specificity = \frac{TN}{TN+FH} \dots\dots\dots (8)$$

$$Specificity = 0.8493(84.93\%)$$

$$4. PPV = \frac{TH}{TH+FH} \dots\dots\dots (9)$$

$$PPV = 0.8533(85.33\%)$$

$$5. NPV = \frac{TN}{TN+FN} \dots\dots\dots (10)$$

$$NPV = 0.8493(84.93\%)$$

Keterangan :

- TH: True Hoax
- TN: True NonHoax
- FH: False Hoax
- FN: False NonHoax
- PPV : Positive Predictive Value
- NPV: Negative Predictive Value

Hasil evaluasi pada penelitian ini menunjukkan bahwa Particle Swarm Optimization dapat meningkatkan nilai akurasi pada Naive Bayes yang merupakan metode yang baik sebagai pengklasifikasi text berita hoax dan non hoax pada media sosial untuk membedakan mana berita yang benar dan mana berita bohong atau hoax.

KESIMPULAN

Pada penelitian ini peneliti menggunakan pengklasifikasi teks berita hoax dan non hoax pada media sosial dengan pengklasifikasi Naive Bayes. Penelitian ini menggunakan 75 data berita hoax dan 75 data berita non hoax pada media sosial. Naive Bayes dapat berfungsi dengan baik sebagai pengklasifikasi text. Akurasi Naive Bayes sebelum

menggunakan Particle Swarm Optimization sebesar 74.67% setelah menggunakan penggabungan metode Naive Bayes dan Particle Swarm Optimization, akurasi menjadi meningkat yaitu menjadi 85.19%. terjadi peningkatan akurasi sebesar 10.52%. Sehingga Naive Bayes berbasis Particle Swarm Optimization pada klasifikasi berita hoax pada media sosial terbukti memberikan nilai akurasi yang lebih akurat.

REFERENSI

friza, A., & Adisantoso, J. (2019). Metode Klasifikasi Rocchio untuk Analisis Hoax, 5(February), 1-10. <http://doi.org/10.29244/jika.5.1.1-10>

Anggono, R., Suryani, A. A., & Kurniati, A. P. (2009). Analisis Perbandingan Metode K-Nearest Neighbor dan Naive Bayes Classifier dalam Klasifikasi Teks. Universitas Telkom.

Detiknews. (2017). Cara Cerdas Mencegah Penyebaran Hoax di Media Sosial. Retrieved from <https://news.detik.com/adv-nhl-detikcom/d-3716300/cara-cerdas-mencegah-penyebaran-hoax-di-media-sosial>

Kurniawan, B., Effendi, S., & Sitompul, O. S. (2012). Klasifikasi Konten Berita Dengan Metode Text Mining. *Jurnal Dunia Teknologi Informasi*, 1(1), 14-19. Retrieved from <http://download.portalgaruda.org/article.php?article=58993&val=4123>

Kussa, P., & Utama, L. (2018). Identifikasi Hoax pada Media Sosial dengan Pendekatan Machine Learning, 13(1), 69-76.

Nurchayati, U. (2019). Ini 11 Macam Berita Bohong yang Perlu Kamu Ketahui. Retrieved from <https://islami.co/ini-11-macam-berita-bohong-yang-perlu-kamu-ketahui/>

Rasywir, E., & Purwarianti, A. (2015). Eksperimen pada Sistem Klasifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin, 3(2), 1-8. Retrieved from <https://www.mendeley.com/import/>

Rozi, F. N., & Sulistyawati, D. H. (2019). Klasifikasi Berita Hoax Pilpres Menggunakan Metode Modified K-Nearest Neighbor Dan Pembobotan Menggunakan Tf-Idf, 15, 1-10.

Wati, R. (2019). *Laporan Akhir Penelitian: Klasifikasi Berita Hoax Pada Media Sosial Menggunakan Algoritma Naive Bayes Dan*



Particle Swarm Optimization. Jakarta.

Yan, P., & Jiao, M. H. (2016). An improved particle swarm optimization for global optimization. *Proceedings of the 28th Chinese Control and Decision Conference, CCDC 2016, 8*, 2181–2185.
<http://doi.org/10.1109/CCDC.2016.7531347>