

# Implementasi Data Mining Pada Pelanggan Telkom Menggunakan Metode K-Nearest Neighbor untuk Memprediksi Status Pelayanan

Ervina Agusetiana<sup>1</sup>, Arif Senja Fitriani<sup>2</sup>

<sup>1,2</sup>Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sidoarjo

E-mail: <sup>1</sup>[161080200234@umsida.ac.id](mailto:161080200234@umsida.ac.id), <sup>2</sup>[asfjim@umsida.ac.id](mailto:asfjim@umsida.ac.id)

**Abstrak** – Penggunaan internet telah menyeluruh di seluruh dunia memudahkan masyarakat untuk mendapatkan sebuah informasi dengan mudah dan cepat. Penelitian ini didasari oleh semakin banyaknya pengguna wifi di Indonesia yang mengakibatkan peningkatan pelanggan wifi PT. Telkom Indonesia. Dampak tersebut mengakibatkan sering terjadi keterlambatan dalam proses pelayanannya. Tujuan penelitian ini ialah untuk mengklasifikasi status pelayanan berupa normal atau tidak normal pada data pelanggan telkom. Penelitian ini mengimplementasikan data mining dengan menggunakan metode k-nearest neighbor. Data yang digunakan ialah data pelanggan Telkom sebanyak 15.113 data pada bulan Mei hingga Juni 2018 dengan dataset sebanyak 100 data. Data diinput ke dalam weka disertai seleksi atribut yang terdiri dari 60% data training dan 40% data testing lalu weka akan memproses perhitungan dan menampilkan hasil klasifikasi yaitu berupa normal ataupun tidak normal. Dari hasil perhitungan klasifikasi pada dataset sebanyak 100 data menggunakan algoritma k-nearest neighbor diperoleh hasil tertinggi menggunakan weka dengan perolehan tingkat kebenaran sebesar 92,5%, tingkat kesalahan sebesar 7,5%.

**Kata Kunci** — data mining; knn; sidoarjo

## 1. PENDAHULUAN

Berdasarkan hasil survei dari Asosiasi Penyelenggara Jasa Internet Indonesia telah didapatkan hasil bahwasanya jumlah pengguna internet di Indonesia sekitar 171,17 juta jiwa dari jumlah keseluruhan penduduk Indonesia yaitu sebesar 264,16 juta jiwa.

Dilihat dari sisi profil pengguna internet di Indonesia pada zaman sekarang ini survei membuktikan, rentang usia 15-19 tahun merupakan kontributor utama dari sisi usia pengguna dengan 91 persen, rentang usia 20-24 tahun sebesar 88,5 persen, rentang usia 25-29 tahun sebesar 82,7 persen, rentang usia 30-34 tahun sebesar 76,5 persen, rentang usia 35-39 tahun sebesar 68,5 persen, rentang usia 40-44 tahun sebesar 51,4 persen, rentang usia 45-49 tahun sebesar 47,6 persen, dan rentang usia 50-54 tahun sebesar 40,9 persen. [1]

Dengan meningkatnya pengguna internet di Indonesia mengakitnya meningkat pula jumlah pelanggan PT. Telkom Indonesia. Namun dari banyaknya *request order* pelanggan telkom, ada beberapa faktor yang mempengaruhi waktu pelayanan dari pihak telkom terhadap pelanggan sehingga status pelayanan ada yang normal dan adakalanya juga tidak normal.

Pada permasalahan di atas, peneliti ingin mengimplementasikan data mining pada data pelanggan telkom menggunakan metode klasifikasi K-Nearest Neighbor untuk memprediksi status pelayanan dan dilaporkan kepada manajer guna mengetahui waktu dalam pelayanan pemasangan wifi sehingga bisa dijadikan evaluasi oleh manajer kepada teknisi agar lebih optimal dalam proses pelayanan.

## 2. METODE PENELITIAN

### 2.1 PT. Telkom

PT Telkom Indonesia ialah sebuah perusahaan yang bergerak pada bidang jasa berupa layanan komunikasi dan jaringan telekomunikasi di Indonesia. Indonesia adalah negara pemegang saham terbanyak yaitu sebesar 52,09% sedangkan 47,91% sisanya dimiliki oleh publik. Saham milik Telkom dengan kode “TLK” kepada New York Exchange (NYSE) dan dengan kode “TLKM” kepada Bursa Efek Indonesia (BEI). [2]

### 2.2 Kualitas Pelayanan

Pelayanan ialah proses pemenuhan kebutuhan yang dilakukan oleh orang lain melalui suatu aktivitas secara langsung. Baik tidaknya pelayanan terletak pada kualitas pelayanan. Kualitas Pelayanan (dalam Tjiptono, 2008:8) merupakan “ukuran seberapa bagus tingkat layanan yang diberikan mampu sesuai dengan ekspektasi pelanggan”. Kualitas pelayanan sangat bergantung pada kemampuan suatu penyedia jasa dalam memenuhi keinginan konsumen yang dilakukan dengan konsisten. Kualitas pelayanan ialah suatu upaya pemenuhan kebutuhan dan keinginan konsumen serta ketepatan penyampaiannya dalam mengimbangi harapan konsumen. [3]

### 2.3 Pelayanan Telkom

Ketersediaan sarana dan prasana yang dimiliki oleh PT. Telkom telah tersebar di seluruh Indonesia. Salah satu contoh akses internet yang dimiliki PT. Telkom ialah Speedy. Speedy yaitu layanan akses internet bagi perumahan yang berkualitas tinggi. PT. Telkom juga memberikan

layanan jaringan wifi. Layanan yang diberikan ialah pasang wifi baru, penambahan maupun kecepatan akses wifi dan pergantian kabel. PT. Telkom mementingkan kepuasan pelanggan dalam pelayanannya. [4]

#### 2.4 Teori Data Mining

Data Mining merupakan proses pengumpulan data yang diolah dengan berbagai macam metode yang sesuai. data mining dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, clustering dan asosiasi (Larose, 2005). Tujuan dari data mining yaitu untuk memanfaatkan data dan mengolahnya hingga mendapatkan sebuah informasi yang baru dan berguna. [5]

#### 2.5 Klasifikasi

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Kamber, 2006). [6]

Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi (Kamber, 2006).

#### 2.6 Metode KNN (K-Nearest Neighbor)

Metode *K-Nearest Neighbor* merupakan salah satu metode yang digunakan dalam pengklasifikasian data. Prinsip kerja *K-Nearest Neighbor* (KNN) adalah mencari jarak terdekat antara data yang akan di evaluasi dengan K tetangga (*neighbor*) terdekatnya dalam data pelatihan (Rismawan, dkk. 2008). Kelebihan dari Algoritma KNN yaitu ketangguhan terhadap training data yang memiliki banyak noise dan efektif apabila training data-nya besar. Sedangkan kelemahan dari algoritma KNN ialah perlu menentukan nilai dari parameter K (jumlah dari tetangga terdekat), training berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan dan atribut mana yang harus digunakan untuk mendapatkan hasil terbaik, dan biaya komputasi cukup tinggi karena diperlukan perhitungan jarak dari tiap query instance pada keseluruhan training sample. [7] Berikut rumus pencarian jarak menggunakan rumus Euclidean (Sutanto, 2009). [8]

$$d_{ij} = \sqrt{\sum_{i=1}^p (x_{ik} - x_{jk})^2} \dots \dots \dots 1$$

Penjelasan dari formula umum algoritma k-nearest neighbor ialah sebagai berikut:

$X_{JK}$  = Sample data/data training  
 $X_{ik}$  = Data uji/ testing  
 $I_j$  = Variabel Data  
 $d$  = Jarak  
 $p$  = Dimensi

#### 2.4 Flowchart

Flowchart adalah sebuah teknik analitis yang digunakan untuk menjelaskan aspek-aspek sistem informasi dengan jelas, tepat dan logis (Krisniaji, 2005). Secara umum, pengertian flowchart merupakan gambaran dari langkah-langkah dan urutan prosedur sebuah program secara grafik. [9]

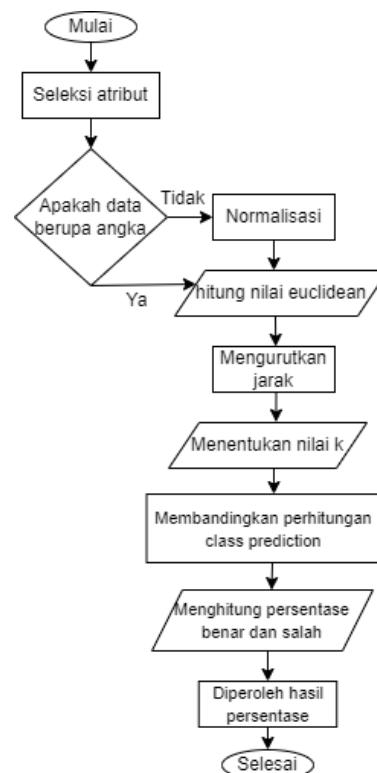
#### 2.5 WEKA

WEKA ialah sebuah paket tools machine learning praktis. “WEKA” merupakan singkatan dari “Waikato Environment for Knowledge Analysis”, yang dibuat di Universitas Waikato, New Zealand untuk penelitian, pendidikan dan berbagai aplikasi. WEKA mengandung tools untuk preprocessing data, klasifikasi, regresi, clustering, aturan asosiasi, dan visualisasi. [10]

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Flowchart

Berikut ini merupakan gambaran flowchart metode klasifikasi algoritma K-NN pada data training dan data testing untuk memprediksi status pelayanan di PT Telkom Indonesia Kab Sidoarjo:



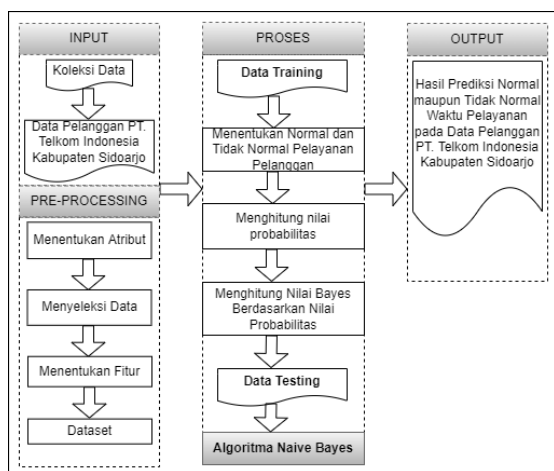
Gambar 3.1 Flowchart Proses K-NN

Dengan adanya gambaran flowchart diatas, Maka tahapan secara umum sesuai klasifikasi algoritma k-nn pada data training sebagai berikut

1. Mulai
2. Menentukan nilai K yang digunakan
3. Menghitung jarak Euclidean
4. Mengurutkan hasil pengurutan jarak
5. Memilih objek yang termasuk dalam nilai K
6. Diperoleh hasil prediksi
7. Selesai

### 3.2 Diagram Klasifikasi

Perancangan diagram klasifikasi dengan menggunakan algoritma knn yaitu sebagai berikut:



Gambar 3.2 Diagram Klasifikasi K-NN

### 3.3 Data Pelanggan Telkom

Data yang digunakan sebagai penelitian ialah data request order dari pelanggan PT. Telkom di Kabupaten Sidoarjo. Data yang akan digunakan untuk memprediksi normal maupun tidaknya waktu pelayanan kepada pelanggan Telkom ialah sebanyak 15.113 data yang diperoleh dari bulan mei sampai dengan juni tahun 2018.

Tabel 3.1 Data Pelanggan Telkom

No	NCLI	NAMA	TC	KC	Status	TT	Tipe	Pelayanan
1.	37284313	AGUS SETYAWAN	ADA	PCT	Complete	AS	1P	N
2.	35835056	AHMAD KHOIRON	TIDAK	DLA	Complete	MO	1P	N
3.	37656944	ABDUL HUSIN	TIDAK	SPJ	UNSCC	AO	3P	TN
4.	37763048	ARIF ARIANTO	ADA	PDA	Process	AO	3P	TN
5.	37724973	NAWI	ADA	SDA	REVOKE	AO	2P	TN
6.	37624509	SITI AMINAH SITI MURTI NENGSIH	ADA	KRN	Complete	AO	3P	N
7.	32438404		TIDAK	TUN	Process	MO	2P	TN
8.	37613772	SITI MUZAYANAH	ADA	SDN	REVOKE	AO	3P	TN
9.	30560673	SITI ROHMAH	ADA	BEJ	Complete	MO	1P	N
10.	34227683	SRI ANIFAH	TIDAK	DLA	Complete	MO	1P	N

### 3.4 Pre-processing

Pada tahap ini dilakukan pre-processing data berupa penentuan atribut dan fitur, seleksi atribut, penentuan dataset, penentuan data training dan penentuan data testing pada data pelanggan Telkom.

### a) Variabel Dan Fitur

Variabel ialah suatu kumpulan yang terdiri dari bagian entitas. Variabel yang digunakan pada data pelanggan Telkom adalah sebanyak 6 atribut dan memiliki class prediksi yaitu pelayanan. Sedangkan fitur merupakan isi dari variabel. Data yang telah diperoleh dari PT. Telkom Indonesia Kabupaten Sidoarjo dapat diketahui variabel dan fitur yaitu sebagai berikut:

### b) Seleksi Atribut

Dari data yang telah diperoleh, telah dilakukan tahap seleksi atribut yang akan diolah melalui 6 atribut, diantaranya ialah variabel TC (Telkom Cabang), variabel KC (Kantor Cabang), variabel Status, variabel Tipe Transaksi (TT), variabel Tipe, variabel Pelayanan dan memiliki class prediksi yaitu Pelayanan. Berikut ini ialah hasil dari pre-processing data pada data request order.

Tabel 3.2 Dataset 100

NO	TC	KC	Status	TT	Tipe	Pelayanan
1.	TIDAK	SDN	Complete	AS	1P	N
2.	ADA	PCT	Complete	AS	1P	N
3.	TIDAK	SDN	Complete	MO	1P	N
4.	TIDAK	DLA	Complete	MO	1P	N
5.	ADA	SDN	Complete	MIGRATE	1P	N
6.	TIDAK	PWS	Complete	MIGRATE	1P	N
7.	TIDAK	SPJ	UNSCC	AO	3P	TN
8.	ADA	SDA	UNSCC	AO	3P	TN
9.	TIDAK	GEM	REVOKE	AO	3P	TN
10.	ADA	SDA	PROCESS	MO	3P	TN
..	..	..	..	..	..	..

Tabel 3.3 Data Training 60

NO	TC	KC	Status	TT	Tipe	Pelayanan
1.	TIDAK	TUN	Complete	MIGRATE	1P	N
2.	TIDAK	TUN	Complete	MIGRATE	1P	N
3.	ADA	SDA	Complete	MO	3P	N
4.	ADA	SDA	Complete	MO	3P	N
5.	ADA	SDA	Complete	MO	3P	N
6.	ADA	SDA	Complete	MO	3P	N
7.	ADA	SDA	Complete	MO	3P	N
8.	ADA	SDA	Complete	MO	3P	N
9.	ADA	GEM	Complete	AS	2P	N
10.	TIDAK	MIP	Complete	MO	3P	N
..	..	..	..	..	..	..

Tabel 3.4 Data Testing 40

NO	TC	KC	Status	TT	Tipe	Pelayanan
1.	ADA	KRN	Complete	MO	3P	?
2.	ADA	SDA	Complete	MO	1P	?
3.	TIDAK	KRN	Complete	MO	2P	?
4.	TIDAK	SPJ	Complete	MO	1P	?
5.	TIDAK	MJS	Complete	MO	3P	?
6.	ADA	PGE	Complete	MO	1P	?
7.	TIDAK	KRN	Complete	MO	2P	?
8.	TIDAK	SDA	REVOKE	AO	3P	?
9.	TIDAK	MRT	Complete	MO	1P	?
10.	TIDAK	PGE	Complete	MO	1P	?
..	..	..	..	..	..	..

### 3.5 Pembahasan

Pada bagian pembahasan akan dijelaskan langkah-langkah dalam melakukan perhitungan menggunakan metode klasifikasi naïve bayes untuk memprediksi normal maupun tidak normal status pelayanan pada data pelanggan Telkom. Data pelanggan tersebut diuji menggunakan tool weka.

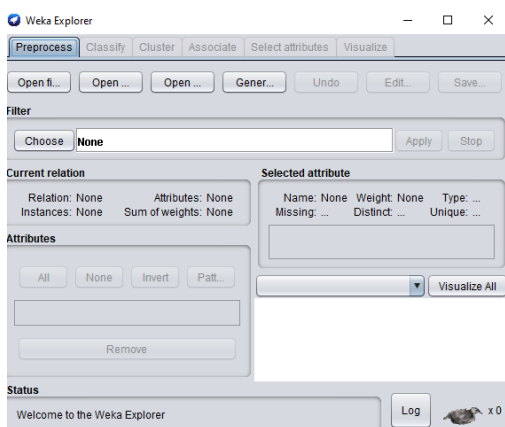
#### 3.5.1 Halaman Awal Weka



Gambar 3.3 Halaman Awal Weka

Halaman ini ialah halaman saat membuka weka. Terdapat 5 menu pada weka diantaranya ialah Explorer, Experimenter, KnowledgeFlow, Workbench dan Simple CLI. Data tersebut akan diuji pada menu Explorer.

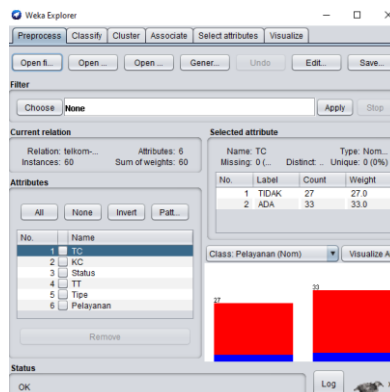
#### 3.5.2 Halaman Awal Explorer



Gambar 3.4 Halaman Menu Explorer

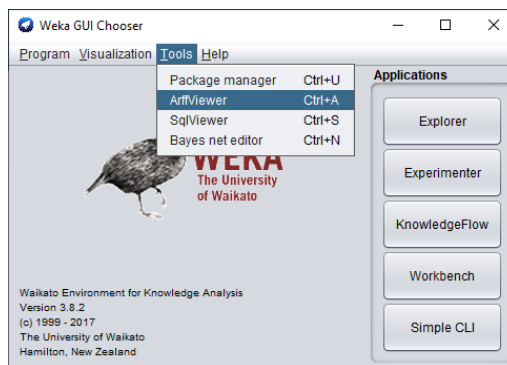
Pada halaman menu explorer terdapat 5 menu pilihan, diantaranya yaitu Preprocess, Classify, Cluster, Associate, Select attribute dan Visualize. Pada tahap pengujian ini akan dilakukan pada menu Preprocess dengan membuka file pada Open File, lalu masukkan data yang akan diuji yaitu data training.

#### 3.5.3 Halaman Atribut Data Training



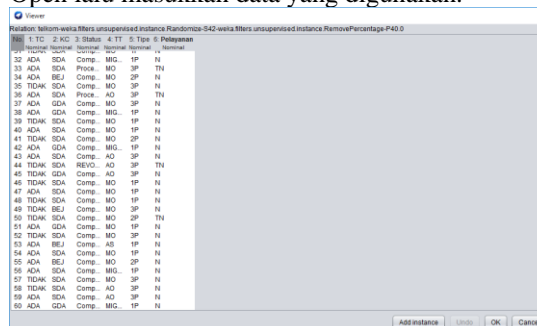
Gambar 3.5 Halaman Atribut Data Training

Setelah memasukkan file, akan muncul atribut yang terdapat dalam data pelanggan. Pada halaman atribut terdapat 6 atribut yang digunakan, diantaranya ialah TC, KC, Status, TT, Tipe dan Pelayanan. Class yang digunakan untuk pengujian yaitu pelayanan dengan nilai ADA sebanyak 33 dan TIDAK sebanyak 27 dari 60 data pelanggan.



Gambar 3.6 Halaman Tools Viewer Atribut

Untuk melihat atribut data training yang telah diubah pada format arff ialah masuk ke halaman awal Weka, masuk ke menu Tools lalu pilih ArffViewer. Pilih File untuk memasukkan data dengan cara klik Open lalu masukkan data yang digunakan.



Gambar 3.7 Halaman Viewer Atribut Data Training

Pada gambar 4.5 menampilkan halaman Viewer yang digunakan untuk melihat atribut data training yang telah diubah pada format arff. Pada halaman tersebut terdapat atribut data training yang digunakan yaitu sebanyak 60 data training.

## 4. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil kesimpulan sebagai berikut:

1. Prediksi klasifikasi status pelayanan pada data pelanggan Telkom dilakukan menggunakan metode klasifikasi KNN serta menggunakan 6 variabel yang telah ditentukan.
2. Hasil prediksi pada data pelanggan Telkom dibagi menjadi 100 dataset. Pada setiap dataset sebanyak 60% dari tiap dataset dijadikan sebagai data training dan sebanyak 40% dari tiap dataset dijadikan sebagai data testing.
3. Dari hasil klasifikasi pada metode KNN untuk memprediksi status pelayanan dari 100 dataset, diperoleh hasil tertinggi menggunakan weka dengan perolehan tingkat kebenaran sebesar 92,5% dan tingkat kesalahan sebesar 7,5%.

## 5. SARAN

1. Pada penelitian berikutnya dapat menggunakan data dari PT. Telkom Indonesia yang didapat dari kabupaten lain.
2. Menggunakan jumlah dataset yang sama untuk mengetahui perbandingan hasil klasifikasi pada metode KNN dengan metode yang berbeda.

## DAFTAR PUSTAKA

- [1] "A. Penyelenggara and J. Internet,," *"RESPONDEN Survei Nasional PENETRASI Pengguna Internet 2018,"*, 2018.
- [2] D. A. Kesawasidhi, *"Pengaruh Kinerja Kualitas Pelayanan Terhadap Kepuasan Dan Loyalitas Pelanggan Pt. Telekomunikasi Indonesia Tbk. Cabang Yogyakarta,"*, pp. Fak. Ekon. Manaj. Univ. Islam Indones., pp. 1–21,, 2017.
- [3] S. A. Wibisono and, *"Pengaruh Kualitas Jasa Pelayanan Terhadap Kepuasan Pelanggan,"*, Vols. vol. 6, no. 2,, pp. "J. Bisnis Akuntansi," 2016.
- [4] I. Oktavianis, *Upaya PT. Telkom Dalam Meningkatkan Kualitas Pelayanan Internet Speedy Dalam Memuaskan Pelanggan (Studi Pada PT. Telkom Malang),*, Vols. vol. 1, no. 1, p. J. Adm. Publik Mhs. Univ. Brawijaya, 2013.
- [5] H. Leidiyana, *Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor*, Vols. vol. 1, no. 1, pp. J. Penelit. Ilmu Komputer, Syst. Embed. Log., 2013.
- [6] S. Sumarlin, *"Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM*, Vols. vol. 5, no. 1, pp. J. Sist. Inf. Bisnis, , 2015.
- [7] W. Yustanti, *"Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah,"*, Vols. vol. 9, no. 1, pp. J. Mat. Stat. dan komputasi,, 2012.
- [8] S. S. a. L. H. S. A. Hasan, *"Klasifikasi Status Gizi Menggunakan K-Nearest Neighbor,"*, Vols. vol. 3, no. 1, 2015.
- [9] gunadharna, *"Definisi dan Simbol Flowchart,"*, p. Defin. Dan Simbol Flowchart, 2016.
- [10] J. H. Y. P. S. F. I. a. I. W. S. W. D. Purnamasari, *Machine Learning 'Get Easy Using WEKA,'*, 2013.