KLASIFIKASI DOKUMEN PUTUSAN PENGADILAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Eri Zuliarso[1], Herny Februarianti[2], Sugiyamto[3]
[1]Program Studi Teknik Informatika, [2,3]Program Studi Sistem Informasi
Fakultas Teknologi Informasi, Universitas Stikubank
e-mail: [1]eri299@edu.unisbank.ac.id, [2]hernyfeb@edu.unisbank.ac.id,
[3]sugiyamtagik@edu.unisbank.ac.id

ABSTRAK

Sistem Informasi Penelusuran Perkara adalah aplikasi berbasis web yang diperuntukan dalam administrasi dan penelusuran terhadap data perkara pengadilan tingkat pertama. layanan informasi berbasis teknologi ini di dalamnya terdapat pencatatan informasi yang sangat lengkap diantaranya . Aplikasi SIPP merupakan bagian dari Sistem Manajemen Informasi di Pengadilan. Hal ini juga bagian dari transparansi terhadap proses peradilan bagi masyarakat umum. Pada metode Support Vector Machine terdapat fungsi kernel yang memetakan data ke ruang vektor yang berdimensi lebih tinggi sehingga kelas dapat dipisahkan secara linear oleh sebuah bidang pemisah. Berdasarkan hasil implementasi dan pengujian yang telah dilakukan pada amar putusan Mahkamah Agung tenang merek dengan menggunakan metode Support Vector Machine ini maka mendapat hasil akurasi sebesar 50%. Dengan hasil akurasi yang cukup tinggi yaitu 50% maka metode Support Vector Machine dapat digunakan untuk melakukan klasifikasi amar putusan kasasi Mahkamah Agung ditolak atau dikabulkan.

Kata Kunci*: Support Vector Machine, klasifikasi, amar putusan*

1. PENDAHULUAN

Sistem Informasi Penelusuran Perkara adalah aplikasi berbasis web yang diperuntukan dalam administrasi dan penelusuran terhadap data perkara pengadilan tingkat pertama sampai dengan tingkat kasasi. Seluruh informasi tersebut dapat diakses oleh publik secara gratis/cuma-cuma dan reatime. Direktori Putusan menampilkan putusan yang dikeluarkan oleh Mahkamah Agung dan pengadilan yang ada di bawahnya. Tujuan 'Direktori Putusan' elektronik ini adalah memastikan adanya informasi yang lengkap bagi masyarakat secara cepat dan murah. Penelusuran putusan Mahkamah Agung yang tersimpan dalam bentuk database pada aplikasi Direktori Putusan ini dapat diakses pencari Informasi, dimana Pencari Informasi dapat memperoleh informasi yang dibutuhkan sehubungan dengan pelayanan kepada masyarakat pada umumnya, sebagai implementasi dari Undang-Undang Republik Indonesia Nomor : 14 Tahun 2008 tentang Keterbukaan Informasi Publik dan SK KMA Nomor: 1-144/KMA/SK/I/2011 tentang Pedoman Pelayanan Informasi di Pengadilan (Mansyur, 2015)

Klasifikasi termasuk teknik pembelajaran mesin atau biasa disebut supervised learning. Menurut Manning et al. (2008), supervised learning adalah proses pembelajaran mengenai ciri dari tiap-tiap kategori yang ada. Teknik ini membangun sebuah classifier yang mempelajari ciri tiap kategori berdasarkan dokumen latih yang dimiliki. Support Vector Machine (SVM) melakukan menganalisis dengan cara belajar dari sekumpulan contoh dokumen yang telah diklasifikasi sebelumnya, menganalisis data dan mengenali pola, digunakan untuk klasifikasi(Cortes & Vapnik, 1995). SVM standar mengambil himpunan data input, dan memprediksi. Untuk setiap masukan yang diberikan, maka ada kemungkinan masukan adalah anggota dari salah satu kelas dari dua kelas yang ada.

Dalam penelitian Liu dkk(2017) melakukan percobaan membandingkan lima model machine learning yang terkenal untuk memprediksi keputusan peradilan: Ã-NN, regresi logistik, bagging, random forests dan SVM. Hasil percobaan menunjukkan bahwa model SVM mengungguli model-model lain dengan pengaturan yang berbeda, dan informasi semantik dalam teks memainkan peran penting dalam memilih fitur untuk model prediksi.

Penelitian Feinerer (2008) menggunakan metode text mining untuk menyelidiki yurisdiksi pengadilan administratif tertinggi Austria mengenai pungutan dan pajak. Model yang dikembangkan menganalisis hukum korporasi untuk melakukan pengelompokan dokumen yurisdiksi ke dalam kelompok pemodelan kelas pajak (seperti pajak penghasilan atau pajak

pertambahan nilai) dan mengidentifikasi properti yurisdiksi. Temuan ini dibandingkan dengan hasil yang diperoleh oleh para ahli hukum. Mahfouz(2010) mengembangkan model Support Vector Machine (SVM) untuk mengekstraksi faktor dasar hukum, yang menjadi dasar hakim mengambil keputusan, baik berdasarkan bukti dari dari kasus-kasus sebelumnya. Model dilatih dan diuji pada dua himpunan kasus berbeda yang dikompilasi dari Pengadilan Federal New York. Dua himpunan masing-masing terdiri dari 120 dan 450 kasus.

Februariyanti (2014) membuat klasifikasi dan visualisasi perundang-undangan berdasarkan dasar hukum. Klasifikasi didasarkan pada kesamaan analisis jejaring yang ada di dasar hukum. Teori graph pada penelitian ini digunakan untuk menampilkan visualisasi dokumen perundang-undangan. Sedangkan dalam makalah ini menggunakan teknik klasifikasi yang mampu memisahkan dokumen secara linear menggunakan metode SVM. Support vector machine merupakan supervised learning di mana metode ini berusaha mencari hyperplane (bidang pemisah) terbaik pada input space, sehingga masalah klasifikasi non-linear dapat diselesaikan dengan cara meningkatkan dimensi ruangnya (Pilászy 2005). Pada metode SVM terdapat fungsi kernel yang memetakan data ke ruang vektor yang berdimensi lebih tinggi sehingga kelas dapat dipisahkan secara linear oleh sebuah bidang pemisah. Dengan demikian metode ini diharapkan dapat menghasilkan kinerja klasifikasi lebih baik dan lebih efisien.

Dalam penelitian ini akan dibahas pemanfaatan metode Support Vector Machine untuk mengklasifikasi putusan pengadilan untuk memudahkan pencarian dokumen putusan pengadilan yang sesuai.

2. METODE PENELITIAN

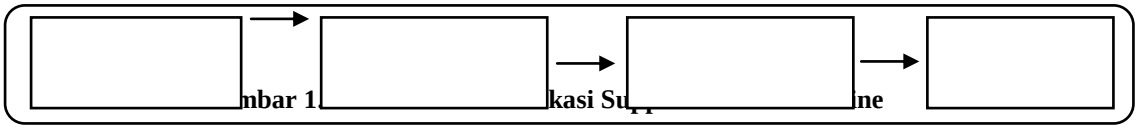
2.1. Obyek Penelitian

Dalam penelitian ini objek yang digunakan adalah *keputusan kasasi Mahkamah Agung* tentang merek. Bagian putusan pengadilan yang akan di klasifikasi adalah bagian amar putusan. Amar putusan (diktum) adalah putusan pengadilan merupakan jawaban terhadap petitum dalam gugatan penggugat.. .

Tabel 1. Contoh Amar putusan MA tentang merek

Kabul	Bahwa Penggugat sangat berkeberatan atas pendaftaran merek Tergugat tersebut di atas karena merek tersebut di atas mempunyai persamaan pada pokoknya dengan merek sudah dikenal & Lukisan milik Penggugat
Kabul	Bahwa ternyata pendaftaran merek dagang Lavera yang dilakukan oleh Tergugat mempunyai persamaan pada pokoknya dengan merek dagang Lavera milik Penggugat yang sudah dikenal di berbagai negara (Pasal 6ayat (la) Undang Undang Nomor 15 tahun 2001 tentang Merek)Kabul
Tolak	Bahwa berdasarkan fakta-fakta yang disampaikan dalam uraian di atas sudah jelaslah bahwa Penggugat adalah Pencipta dan Pemegang Hak Cipta atas Seni Motif Kertas Nomor 020607
Tolak	Eksepsi Tentang Penggugat tidak mempunyai hak dan kapasitas untuk menggugat (non legitima persona standi in judicio) Karena Penggugat tidak ada hubungan hukum dengan Tergugat atas merek dan Penggugat a quo tidak mempunyai hak tunggal/khusus untuk dan atas merek tersebut

Metode analisis data merupakan proses yang akan dilakukan untuk menganalisa data amar *putusan kasasi Mahkamah Agung* sehingga didapat kelas apakah diterima atau di kabulkan. Berikut adalah gambaran penyelesaian masalah dalam penelitian ini dapat dilihat pada gambar 1.



mbar 1. kasi Su ne

Berikut adalah gambaran penyelesaian klasifikasi putusan kasasi Mahkamah Agung pada gambar 2

1. Analisis Sumber Data
   Setelah data didapatkan, langkah selanjutnya ialah melakukan penentuan sentimen secara manual terhadap data tersebut. Data yang diambil berjumlah 14 data klasifikasi amar putusan kasasi Mahkamah Agung terdiri dari 10 data latih dan 4 data test. Untuk data latih terdiri dari 5 klasifikasi putusan kasasi Mahkamah Agung dengan keputusan ditolak dan 5 klasifikasi putusan kasasi Mahkamah Agung dengan keputusan dikabulkan.
2. Data *Preprocessing*
   Pada tahap ini, klasifikasi amar putusan kasasi Mahkamah Agung yang telah dikumpulkan akan melalui beberapa tahap pemrosesan teks yang terdiri dari *case folding*, *tokenizing* dan *stopword removal*. Tujuan dari pemrosesan teks adalah supaya data yang didapat akan lebih terstruktur agar lebih mudah untuk dilakukan pengolahan data.

3. Klasifikasi putusan kasasi Mahkamah Agung Support Vector Machine
   Tahap selanjutnya adalah proses pengklasifikasian yang akan diproses menggunakan metode algoritma Support Vector Machine dengan menghitung probabilitas dari tiap putusan kasasi Mahkamah Agung untuk menentukan data mana yang termasuk dalam keputusan yang dikabulkan dan ditolak.
4. Hitung Akurasi
   Data yang diuji adalah data yang telah diketahui sentimennya, sehingga setelah dilaukan klasifikasi sentimen maka akan dilakukan perhitungan akurasi atau ketepatan sistem dalam melakukan klasifikasi. Dalam pengukuran akurasi metode yang digunakan dalam penelitian ini adalah *Confusion Matrix*.

Untuk dapat mengklasifikasikan *Putusan kasasi* ke dalam kategori yang telah ditentukan maka data *Putusan kasasi* harus melalui beberapa proses. Pemrosesan klasifikasi diperlukan sebuah *tools*, dalam penelitian ini *tools* yang digunakan adalah Bahasa R.

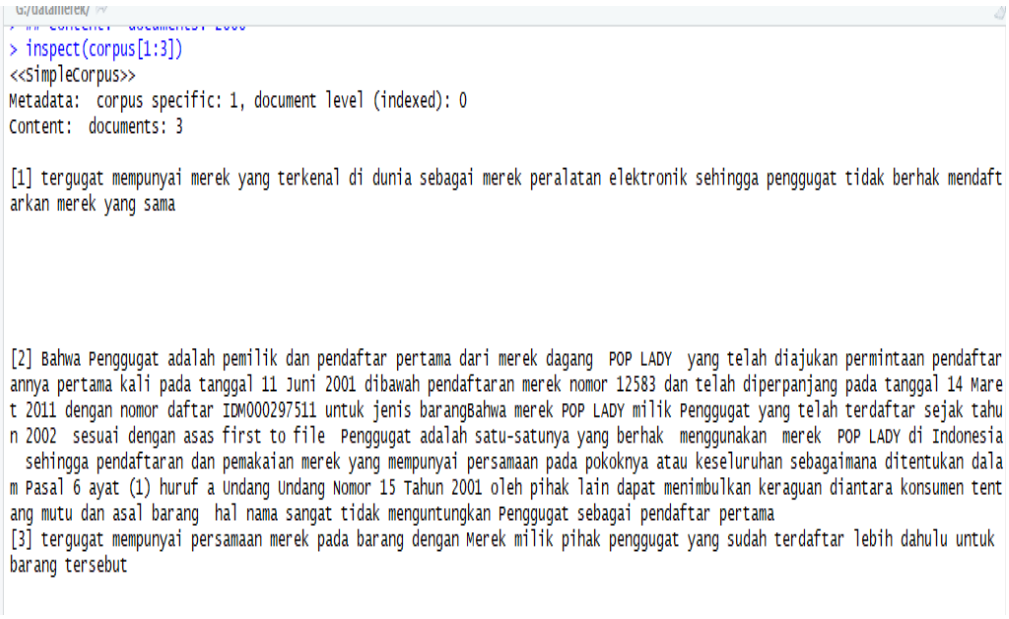
```
> df <- read_excel( G:/datamerek/datamerek.xlsx )
> glimpse(df)
Observations: 14
Variables: 2
$ label <chr> "kabul", "kabul", "kabul", "kabul", "kabul", "tolak", "tolak", "tolak", "tolak", "tolak", "tola...
$ text  <chr> "karena merek bodycology terdaftar No. IDM000289450 atas nama Tergugat secara jelas mempunyai p...
>
```

Gambar 2. Data frame hasil membaca data amar putusan MA

Gambar 2 adalah tampilan dari hasil membaca data *datamerek.xlsx* yang berisi 14 data *putusan kasasi Mahkamah Agung* dengan 2 variabel yaitu *label dan text*. Variabel *label* berisi label kelas pada putusan kasasi Mahkamah Agung , label kelas “tolak” mewakili kelas gugatan ditolak Mahkamah Agung. Sedangkan variabel *text* berisi kumpulan amar *putusan kasasi Mahkamah Agung* .

2.2. Membuat Corpus

Dalam pendekatan ini, *corpus* mewakili setiap kata dalam dokumen sebagai token atau fitur dan setiap dokumen sebagai vektor fitur. Disini urutan kata diabaikan dan fokus hanya pada jumlah kemunculan setiap kata. Data yang digunakan untuk membuat corpus adalah data *datamerek.xlsx* dengan variabel text. Sedangkan *df$label* digunakan untuk menentukan sebagai label klasifikasi apakah keputusan ditolak atau dikabulkan



```
> inspect(corpus[1:3])
<<SimpleCorpus>>
Metadata:  corpus specific: 1, document level (indexed): 0
Content:  documents: 3

[1] tergugat mempunyai merek yang terkenal di dunia sebagai merek peralatan elektronik sehingga penggugat tidak berhak mendaft
arkan merek yang sama

[2] Bahwa Penggugat adalah pemilik dan pendaftar pertama dari merek dagang  POP LADY  yang telah diajukan permintaan pendaftar
annya pertama kali pada tanggal 11 Juni 2001 dibawah pendaftaran merek nomor 12583 dan telah diperpanjang pada tanggal 14 Mare
t 2011 dengan nomor daftar IDM000297511 untuk jenis barangBahwa merek POP LADY milik Penggugat yang telah terdaftar sejak tahu
n 2002  sesuai dengan asas first to file  Penggugat adalah satu-satunya yang berhak  menggunakan  merek  POP LADY di Indonesia
  sehingga pendaftaran dan pemakaian merek yang mempunyai persamaan pada pokoknya atau keseluruhan sebagaimana ditentukan dala
m Pasal 6 ayat (1) huruf a Undang Undang Nomor 15 Tahun 2001 oleh pihak lain dapat menimbulkan keraguan diantara konsumen tent
ang mutu dan asal barang  hal nama sangat tidak menguntungkan Penggugat sebagai pendaftar pertama
[3] tergugat mempunyai persamaan merek pada barang dengan Merek milik pihak penggugat yang sudah terdaftar lebih dahulu untuk
barang tersebut
```

Gambar 3. Hasil *Corpus*

2.3. *Text Processing*

Tindakan yang dilakukan pada tahap ini adalah tokenisasi, mengubah semua huruf menjadi huruf kecil, menghapus tanda baca, menghapus angka dan menghapus kata – kata (*stopword removal*) yang tidak memiliki makna sehingga tidak berpengaruh terhadap proses analisis sentimen. *Stopword* yang digunakan merupakan modifikasi *stopword* (*custom stopword*). Penggunaan *custom* stopword karena dalam RStudio belum tersedianya *stopword* untuk bahasa Indonesia.

custom_stopwords memiliki 796 daftar kata. Daftar kata inilah yang akan dihilangkan jika ditemukan dalam *putusan kasasi Mahkamah Agung* . Untuk menjalankan tahap *text processing* script yang digunakan gambar 4.

```
corpus.clean <- corpus %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removePunctuation) %>%
  tm_map(removeNumbers) %>%
  tm_map(removeWords, custom_stopwords) %>%
  tm_map(stripWhitespace)
```

Gambar 4. *Script Text Processing*

2.4 Document Term Matrix (DTM)

Document Term Matrix (DTM) adalah tahapan mempresentasikan kata-kata (*term*) menjadi matriks. Baris DTM sesuai dengan dokumen dalam dataset, kolom-kolom yang cocok dengan *term*, dan itu adalah unsur-unsur frekuensi *term*. Untuk membangun DTM menggunakan fungsi built-in dari *packages* ‘tm’.

Gambar 5 adalah contoh dari beberapa *term* yang telah dibuat matriks. Data yang diambil berjumlah 3 dokumen dengan 10 *term*.

```
> dtm.test <- dtm[11:14,]
> inspect(dtm.train[1:3,1:10])
<<DocumentTermMatrix (documents: 3, terms: 10)>>
Non-/sparse entries: 18/12
Sparsity           : 40%
Maximal term length: 12
Weighting          : term frequency (tf)
Sample             :
    Terms
Docs berhak dunia elektronik mempunyai mendaftarkan merek penggugat peralatan sama sebagai
   1      1     1         1         1            1     3         1         1    1       1
   2      1     0         0         1            0     5         4         0    0       1
   3      0     0         0         1            0     2         1         0    0       0
>
```

Gambar 5. Contoh *Document Term Matrix*

3. HASIL DAN PEMBAHASAN

3.1. Hasil dan Pembahasan

Data *putusan kasasi Mahkamah Agung* akan dibagi menjadi data latih dan data uji. Dari 14 data putusan kasasi Mahkamah Agung , data latih yang dipakai berjumlah 10 data *putusan kasasi Mahkamah Agung* dan data uji berjumlah 4 data *putusan kasasi Mahkamah Agung*. Data latih terdiri dari 5 data *putusan kasasi Mahkamah Agung* dengan keputusan ditolak dan 5 data *putusan kasasi Mahkamah Agung* dengan keputusan kabul. Data latih menghasilkan 2.310 *term.*. Setelah dilakukan *preprocessing*, *term* pada data latih berkurang menjadi 1.831 *term*.

3.2 Klasifikasi *Support Vector Machine*

. Konsep Support Vector Machine (SVM) secara sederhana merupakan usaha mencari Hyperplane terbaik yang berfungsi sebagai pemisah dua buah class pada input space. Untuk n-dimensional space, input data xi (i=1. . .k), dimana milik kelas 1 atau kelas 2 dan label yang terkait menjadi -1 untuk kelas 1 dan +1 untuk kelas 2..

```
G:/datamerek/
> # print a summary
> fit1

Call:
svm(formula = cat ~ ., data = sms_train1)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1
      gamma:  0.009433962

Number of Support Vectors:  10

>
> fit1.pred <- predict(fit1, na.omit(sms_test1))
> fit1.pred
    1     2     3     4
kabul tolak kabul kabul
Levels: kabul tolak
```

Gambar 6. Hasil Prediksi Data Uji

Gambar 6 merupakan hasil prediksi data uji menggunakan algoritma *Support VectorMachine*.

3.3 *Confusion Matrix* (Menghitung Akurasi)

Dari hasil prediksi klasifikasi kelas putusan kasasi Mahkamah Agung data uji pada tahap ini akan dihitung akurasi dari algoritma *Support Vector Machine* dalam melakukan klasifikasi. Data hasil klasifikasi algoritma *Support Vector Machine* akan dicocokan dengan kelas putusan kasasi Mahkamah Agung yang telah diketahui sebelumnya.

```
> confMatrix1
Confusion Matrix and Statistics

          Reference
Prediction kabul tolak
     kabul     1     2
     tolak     0     1
```

Gambar 7. Hasil *Confusion Matrix*

Dari gambar 7 hasil pencocokan data sentimen klasifikasi algoritma *naive bayes* dengan sentimen sebenarnya menghasilkan :

1. *True* Positif 1 dokumen
2. *True* Negatif 2 dokumen
3. *False* Positif 0 dokumen
4. *False* Negatif 1 dokumen

```
> confMatrix1
Confusion Matrix and Statistics

          Reference
Prediction kabul tolak
     kabul     1     2
     tolak     0     1

               Accuracy : 0.5
                 95% CI : (0.0676, 0.9324)
    No Information Rate : 0.75
    P-Value [Acc > NIR] : 0.9492

                  Kappa : 0.2
 Mcnemar's Test P-Value : 0.4795

            Sensitivity : 1.0000
            Specificity : 0.3333
         Pos Pred Value : 0.3333
         Neg Pred Value : 1.0000
             Prevalence : 0.2500
         Detection Rate : 0.2500
   Detection Prevalence : 0.7500
      Balanced Accuracy : 0.6667

       'Positive' Class : kabul
```

Gambar 8. Akurasi Klasifikasi Algortima *Naive Bayes*

Pada Gambar 8 akurasi dari algoritma *Support Vector Machine* dalam melakukan klasifikasi *putusan kasasi Mahkamah Agung* dalam penelitian ini sebesar 0.5 = 50%.

4. KESIMPULAN

Berdasarkan hasil implementasi dan pengujian yang telah dilakukan pada sistem analisis sentimen dengan menggunakan metode Support Vector Machine ini maka dapat diambil kesimpulan sebagai berikut :

1. Metode Support Vector Machine dalam melakukan klasifikasi putusan kasasi Mahkamah Agung ditolak dan dikabulkan dengan 10 data latih dan 4 data uji mendapat hasil akurasi sebesar 50%.
2. Dengan hasil akurasi yang cukup tinggi yaitu 50% maka metode Support Vector Machine perlu ditingkatkan untuk dapat digunakan untuk melakukan klasifikasi putusan kasasi Mahkamah Agung ditolak dan dikabulkan.

DAFTAR PUSTAKA

[1] Cortes, C. & Vapnik, V. 1995. Support-Vector Networks Machine Learning, 20. (Online). http://www.springerlink.com/content/k238jx0 4hm87j80g/.

[2] Feinerer I., Hornik K. ,2008, Text Mining of Supreme Administrative Court Jurisdictions. In: *Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R. (eds) Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin, Heidelberg

[3] Februariyanti, Herny, Zuliarso, Eri,2014, Visualisasi Keterkaitan Perundang-Undangan Di Indonesia,Seminar Nasional Teknologi Industri 2014,Fakultas Teknologi Industri, Universitas Trisakti, Jakarta

[4] Liu, Zhenyu , Huanhuan Chen,2017, *A predictive performance comparison of machine learning models for judicial cases*, IEEE Symposium Series on Computational Intelligence (SSCI)

[5] Mahfouz, Tarek ,Kandil, Amr, 2010, Construction Legal Decision Support Using Support Vector Machine (SVM), *Construction Research Congress 2010, Innovation for Reshaping Construction Practice*,Banff Alberta, Canada

[6] Manning CD, Raghavan P, Schütze H. 2008. An Introduction to Information Retrieval. Cambridge (GB): Cambridge Univ Pr.

[7] Mansyur,R., 2015, Keterbukaan Informasi Di Peradilan Dalam Rangka Implementasi Integritas Dan Kepastian Hukum, Jurnal Hukum dan Peradilan, Volume 4, Nomor 1 Maret 2015 : 83-100

[8] Pilászy I. 2005. Text categorization and support vector machines. Di dalam: The Proceedings of the 6th International Symposium of Hungarian Researchers on Computational Intelligence, 2005.