

# Introduction to ML

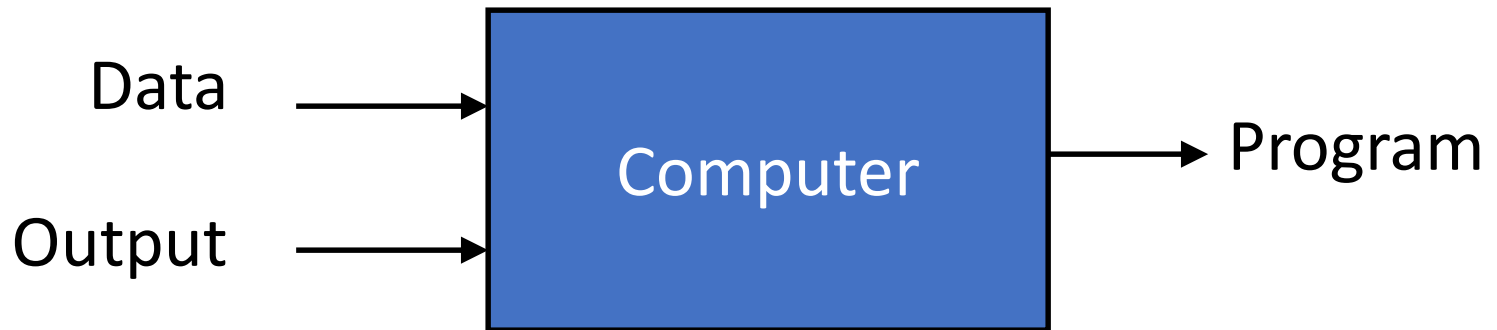
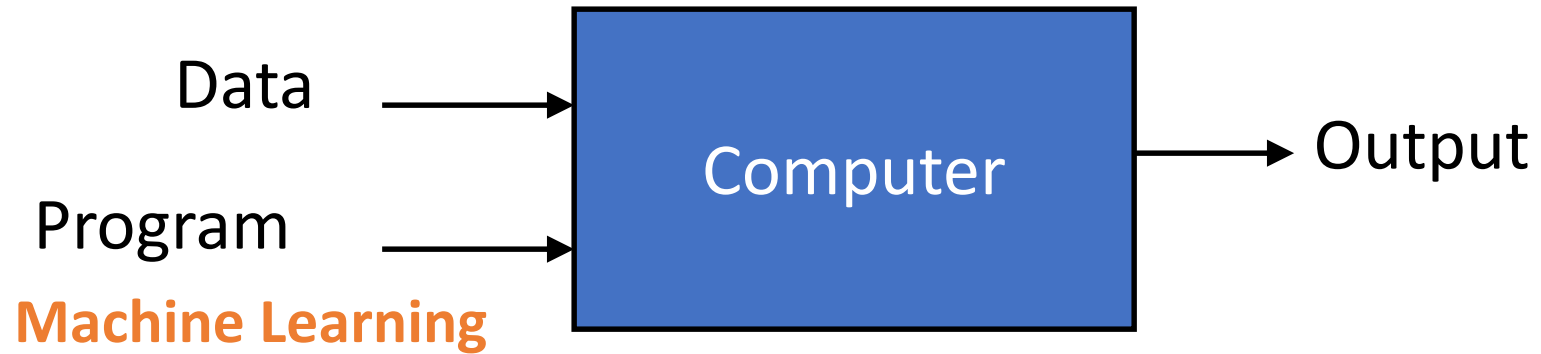
MOF.JS – 13 Februari 2019

**Bakhtiar Amaludin - [github.com/maziyank](https://github.com/maziyank)**

# Machine Learning?

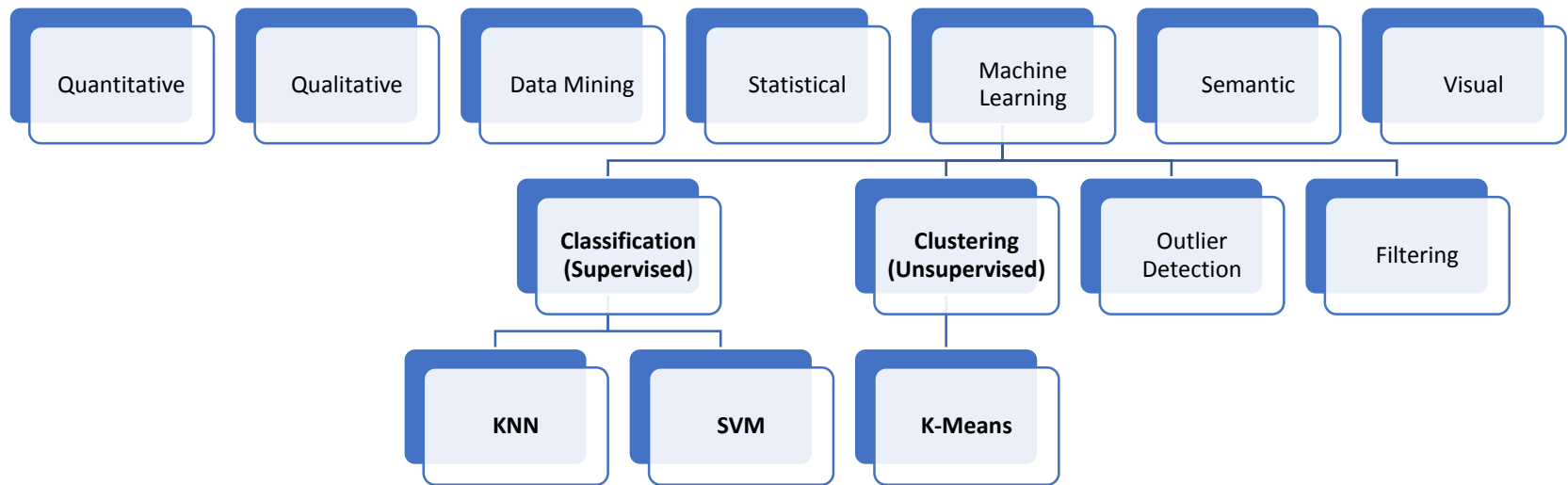
**Machine learning focuses on the development of computer programs** that can access data and use it learn for themselves and later improve from experience without being explicitly programmed.

## Traditional Programming

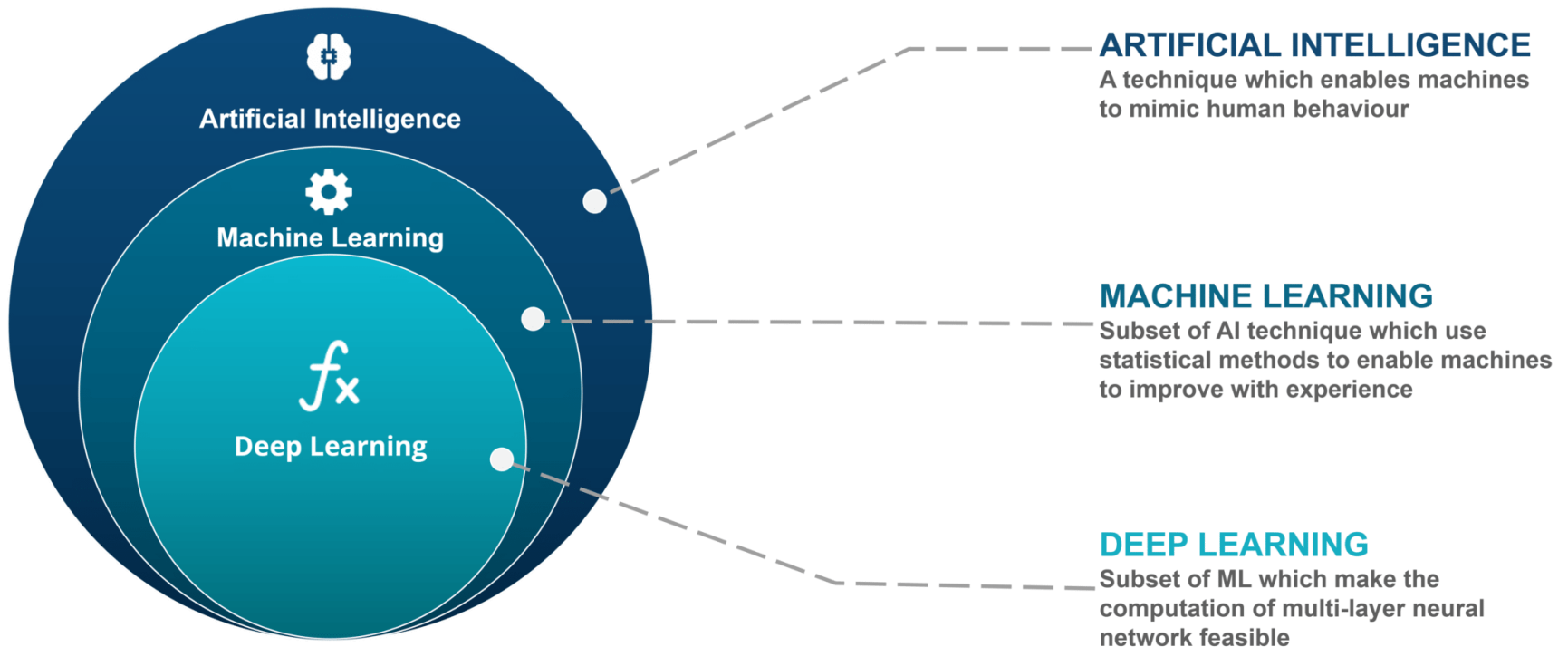


# ML & Big Data

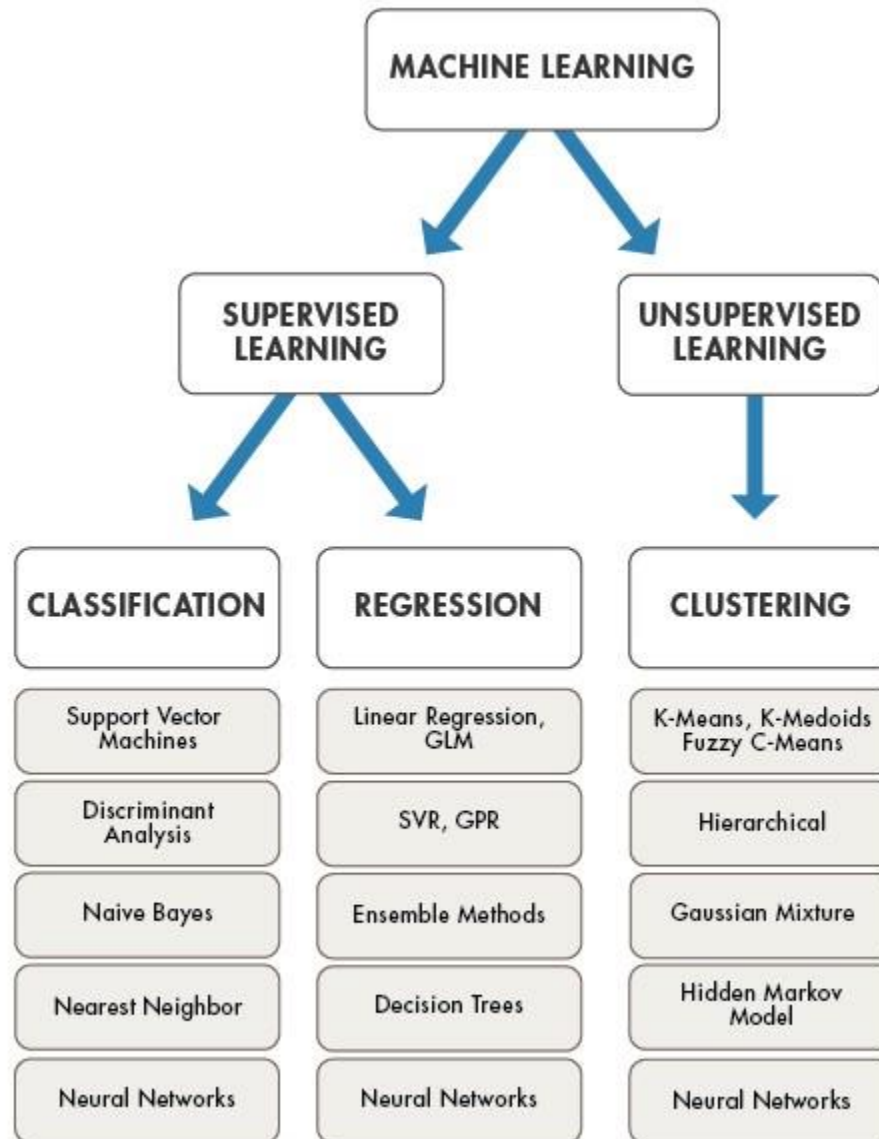
## Big Data Analysis Technique (Erl et al., 2016)



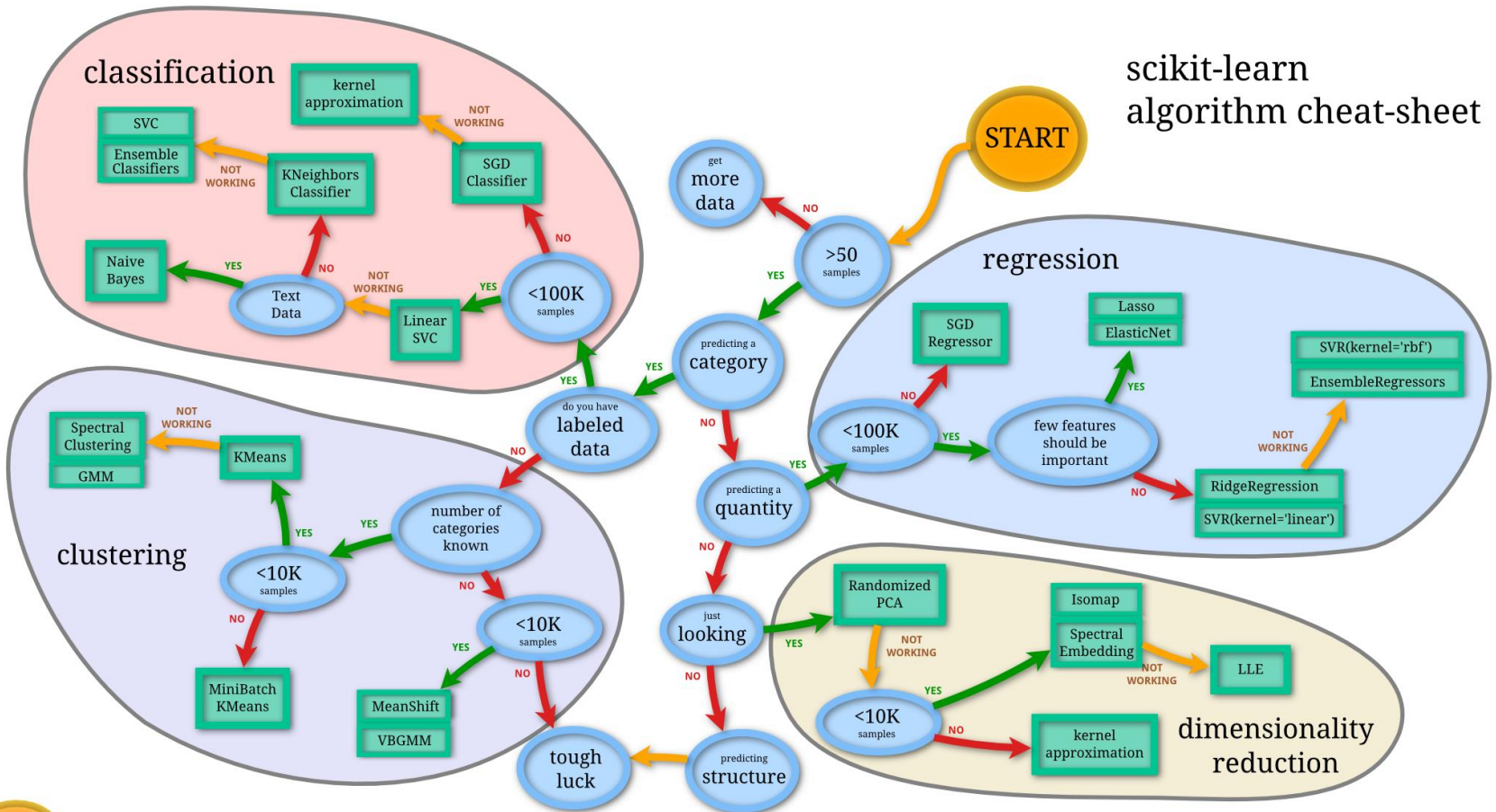
# AI, ML & DL

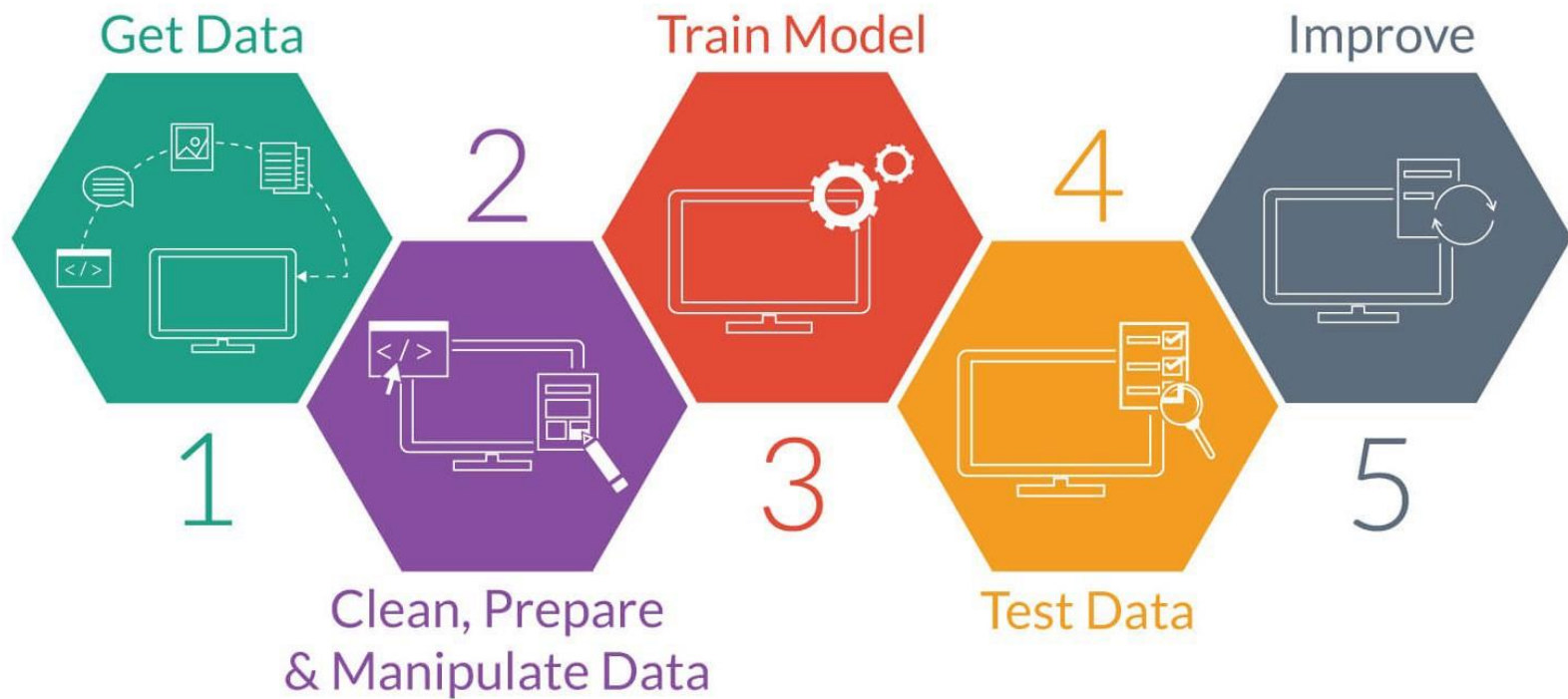


# Machine Learning Algorithm



# scikit-learn algorithm cheat-sheet





# Case Study: Very Simple Document Classification

Goal: Classifying document based on title

judul	EselonIII
Prosedur Pengajuan Dan Penyelesaian Permohonan Pembetulan Keteta	2
Penetapan Tarif Bea Masuk atas impor Kacang Kedelai	2
Penyesuaian besarnya peredaran bruto bagi wajib pajak orang pribadi ya	2
Perkiraan Alokasi Dana Bagi Hasil Sumber Daya Alam Pertambangan Pan	3
Biaya Promosi Yang Dapat Dikurangkan Dari Penghasilan Bruto.	2
Pedoman Akuntansi Dan Pelaporan Aset Berupa Barang Milik Negara Yar	3
Penilaian Barang Milik Negara.	4
Perkiraan Alokasi Dana Bagi Hasil Sumber Daya Alam Kehutanan Tahun A	3
Tata Cara Pengelolaan Dan Pertanggungjawaban Fasilitas Dana Geotherm	2
Tata cara penyediaan, penghitungan dan pembayaran subsidi Bahan Bak	3
Pengelolaan Kelebihan/Kekurangan Kas Pemerintah.	3
Perpanjangan masa berlakunya Keputusan Menteri Keuangan Nomor 11.	2
Dana Operasional Menteri / Pejabat setingkat Menteri.	3
Pengelolaan Barang Milik Negara Yang Berasal Dari Barang Rampasan Ne	4
Penetapan Perkiraan alokasi dana bagi hasil Pajak Bumi dan Bangunan di	3
Pembebasan Bea Masuk Atas Impor Cold Rolled Coil (CRC).	2
Perpanjangan Penetapan Tarip Bea Masuk dalam rangka Normal Track AS	2
Perubahan Kedua Atas Peraturan Menteri Keuangan Nomor 79/PMK.010	3
Pengelolaan Dana Reboisasi Dalam Rekening Pembangunan Hutan.	3
Pengelolaan Kekayaan Negara yang berasal dari BPPN oleh PT Perusahaan	3

# Data Preprocessing

- Split Data for Training and Testing
- Dedupe
- Tokenizer
- Stemming
- Remove Stopword
- Vectorization

# Model Training

- Naive Bayes

```
In [16]: from sklearn.naive_bayes import MultinomialNB, BernoulliNB  
nb = MultinomialNB(alpha=0.2)
```

```
In [17]: %time nb.fit(X_train_dtm, y_train)
```

Wall time: 227 ms

```
Out[17]: MultinomialNB(alpha=0.2, class_prior=None, fit_prior=True)
```

- SVM

```
▶ In [23]: from sklearn import svm  
from sklearn.metrics import accuracy_score
```

```
clf = svm.SVC(kernel='linear')  
print("Training Classifier...")  
%time clf.fit(X_train_dtm, y_train)
```

Training Classifier...

Wall time: 405 ms

```
Out[23]: SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,  
decision_function_shape='ovr', degree=3, gamma='auto', kernel='linear',  
max_iter=-1, probability=False, random_state=None, shrinking=True,  
tol=0.001, verbose=False)
```

# Testing

	judul	target	prediction	correct
2062	Pemberian Premi.	2.0	2.0	True
2939	Deklarasi Inisiatif (Voluntary Declaration) At...	2.0	2.0	True
989	Pemberian Fasilitas Pengurangan Pajak Penghasi...	6.0	3.0	False
2237	Kriteria dan/atau Rincian Ternak, Bahan Pakan ...	6.0	6.0	True
490	Mekanisme Pengelolaan Hibah Millennium Challen...	3.0	3.0	True
510	Perubahan Kedua atas Keputusan Menteri Keuanga...	2.0	2.0	True
3057	Pengangkatan Dalam Jabatan Struktural Melalui ...	5.0	5.0	True
3247	Penyusunan, Penetapan, Dan Penerapan Standar P...	5.0	3.0	False
3245	Perubahan Atas Peraturan Menteri Keuangan Nomo...	4.0	4.0	True
4426	Tarif Layanan Badan Layanan Umum Politeknik Ke...	3.0	3.0	True
3512	Tata Cara Penyediaan, Pencairan, dan Pertanggu...	3.0	3.0	True
1837	Penetapan Barang Ekspor Yang dikenakan Bea Kel...	2.0	2.0	True
2049	Pajak Pertambahan Nilai ditanggung pemerintah ...	2.0	2.0	True
1759	Perubahan Atas Peraturan Menteri Keuangan Nomo...	3.0	3.0	True

## Model Accuracy

0.865

Naive Bayes

0.898

SVM

# Case Study: Document Classification

## Machine Learning

This is a simple demo of Document Classifying Using Machine Learning.

Input

Pajak Penghasilan Badan

Prediction

Bagian Hukum Pajak dan Kepabeanan ( ID : 2)

Reset

Predict

Mari Kita Diskusikan.

*Potential ML implementation  
in Gov Serctor?*