

Validator Situng Pilpres 2019 (VSP-2019)

adalah validator hasil hitung suara Pemilu Presiden & Wakil Presiden RI 2019 pada website pemilu2019.kpu.go.id berdasarkan form C1 yang terlampir. Pembuatan aplikasi yang berbasis Python 3.6 ini dilatarbelakangi tingginya kesalahan input oleh pihak KPU sehingga menurunkan tingkat kepercayaan masyarakat kepada instansi Pemerintah. Oleh karena itu, diperlukan validator yang bekerja instan dan akurat dalam memvalidasi data yang masuk.

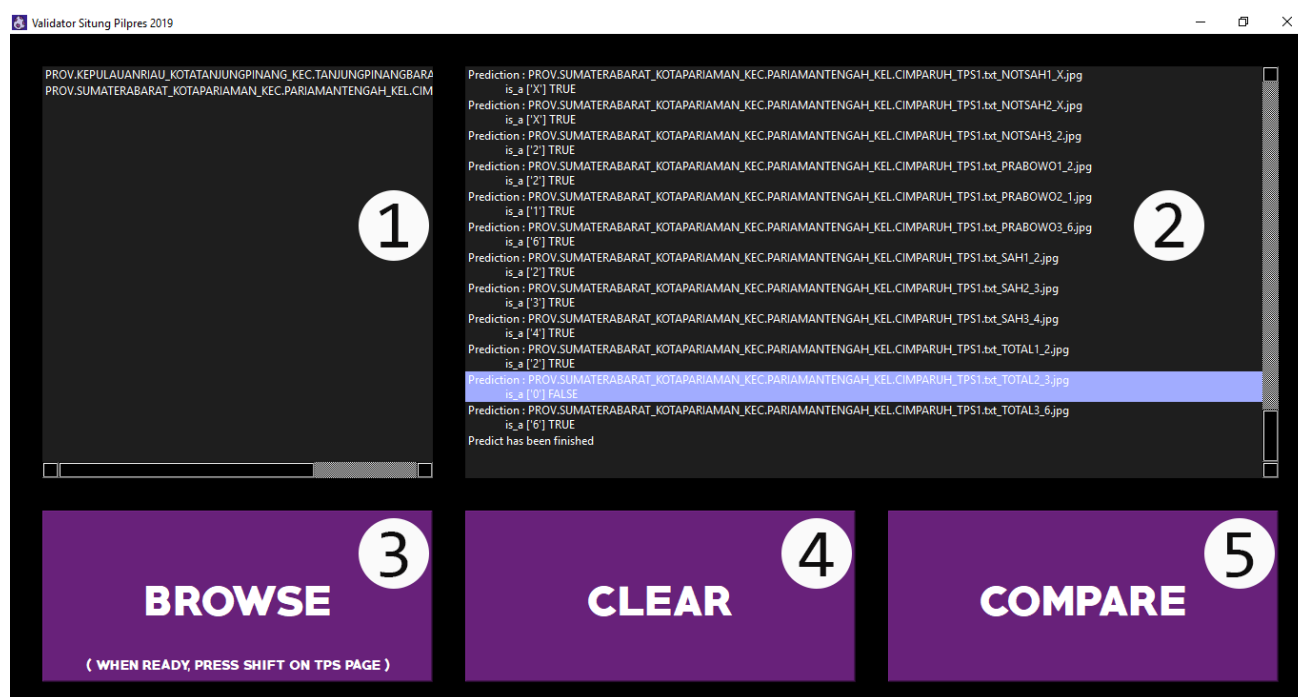
Creator

2101690774 - Yukha Dharmeswara

2101696645 – Calvin Hartanto Honggare

Penggunaan Aplikasi

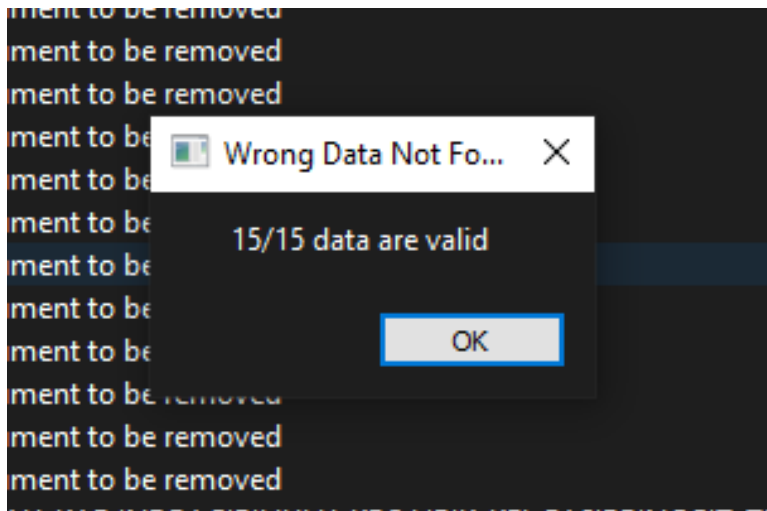
Penggunaan VSP-2019 sangatlah mudah, anda perlu mengklik tombol Browse untuk memilih TPS yang ingin dicek, jika anda sudah berada pada halaman TPS yang diinginkan, tekan tombol “Shift” pada Keyboard anda. Jika data TPS telah dimuat, anda tinggal mengklik tombol Compare untuk melaksanakan proses validasi. Aplikasi akan memunculkan hasil validasi pada Kolom Log (elemen no. 2 pada gambar di bawah).



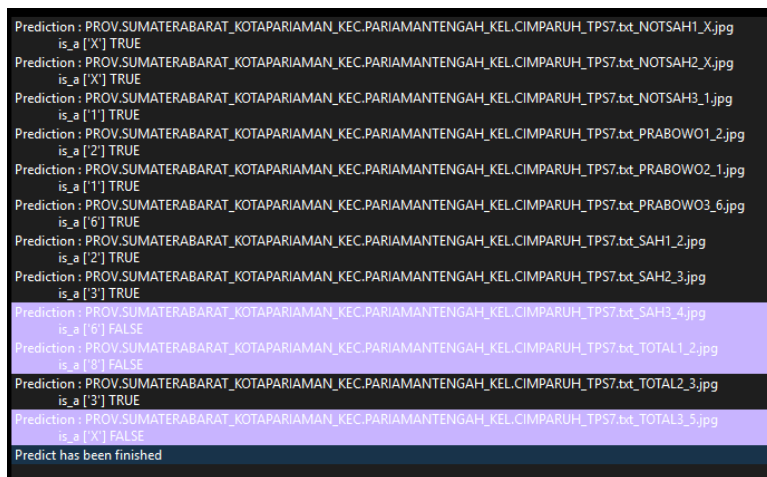
Legenda :

1. Kolom TPS : berisi list TPS yang telah dimuat.
2. Kolom Log : berisi log dan hasil validasi.
3. Tombol Browse : untuk memilih TPS yang ingin divalidasi.
4. Tombol Clear : untuk menghapus data TPS yang telah dimuat.
5. Tombol Compare : untuk melaksanakan validasi.

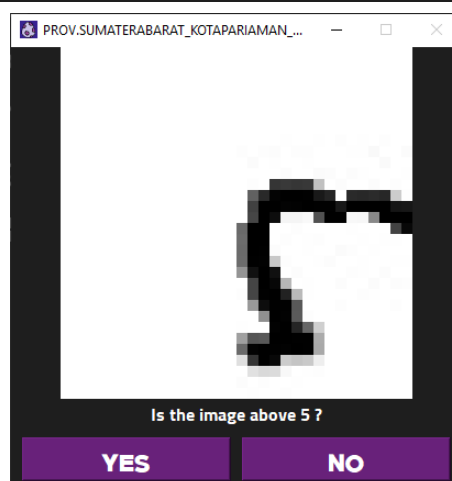
Jika setiap data valid, akan muncul pop-up seperti berikut.



Namun, jika terdapat data yang terprediksi invalid, akan di highlight pada Kolom Log. User dapat mengklik setiap hasil prediksi pada Kolom Log untuk mengecek secara manual.



Misprediksi



Berikut adalah contoh misprediksi yang dilakukan oleh VSP-2019. Website KPU menampilkan angka 5, namun aplikasi memprediksi kelas pada gambar adalah X. Jika kita mengklik log tersebut, dapat dengan jelas angka yang muncul adalah angka 5. Untuk mengatasinya, user dapat mengklik tombol “YES” untuk mengkoreksi label yang diberikan. Namun jika angka yang muncul berbeda dari yang ada di Website KPU, user dapat mengklik tombol “NO” untuk melihat form C1 untuk analisa lebih lanjut.

PROV. RIAU_KAB. INDRAGIRIHULU_KEC. LIRIK_KEL. PASIRRINGGIT_TP55_X02.jpg

Nomor TPS : 5
 Kecamatan / Distrik : LIRIK
 Provinsi : RIAU

Desa / Kelurahan : PASIR RINGGIT
 Kabupaten / Kota : INHU

MODEL C1-PPWP
 Lembar 2

IV. DATA PEROLEHAN SUARA PASANGAN CALON PRESIDEN DAN WAKIL PRESIDEN

NOMOR DAN NAMA PASANGAN CALON		SUARA SAH		
(1)	(2)	(3)		
01.	Ir. H. JOKO WIDODO Prof. Dr.(H.C.) KH. MA'RUF AMIN SERATUS DATU	1	0	1
02.	H. PRABOWO SUBIANTO H. SANDIAGA SALAHUDDIN UNO SERATUS TLGA PULUH DELAPAN	1	3	8

V. DATA SUARA SAH DAN TIDAK SAH

URAIAN	JUMLAH
--------	--------

Cara Kerja Aplikasi

1. Men-scrape data dari website KPU.
2. Mengekstrak region of interest dari kolom angka pada form C1.
3. Memprediksi angka yang muncul pada region of interest menggunakan LinearSVC dengan fitur HoG.
4. Membandingkan angka dari website dengan dari hasil ekstrak C1.

Training

• Dataset

Dataset didapat dari sumber website pemilu2019.kpu.go.id, khususnya dari Kabupaten Aceh Barat, Kabupaten Aceh Barat Daya, dan Kabupaten Aceh Besar pada tanggal 11 Mei 2019. Dataset ini terdiri dari 16319 digit (14120 training data + 2199 testing data) berukuran 32x32 piksel. Dataset dibagi menjadi 11 kelas, yaitu 0,1,2,3,4,5,6,7,8,9, dan X. Karena ada beberapa faktor yang menyebabkan dataset tidak dapat digunakan (Salah input, salah form, pemberian tanda X tidak formal, dicoret-coret, digit ekstraktor gagal mengekstrak data), diperlukan pembersihan secara manual, yang semestinya 23190 digit menjadi 16319 digit.

• Feature

Fitur yang dipakai untuk mengklasifikasi digit adalah Histogram of Oriented Gradients. Histogram of Oriented Gradients adalah feature descriptor berdasarkan histogram/distribusi dari orientasi gradien

local pada gambar. HoG berguna karena dapat mengenali edge dan corner pada gambar yang merupakan bagian vital pada object recognition. HoG diperoleh dengan cara membagi gambar ke dalam beberapa cell, tiap cell akan dihitung orientasi gradiennya, lalu akan di sederhanakan menjadi n-bin(default skimage.feature.hog 9-bin).

Cell akan dikelompokkan menjadi beberapa block untuk dinormalisasi. Secara default, skimage.feature.hog menggunakan L2-Hys, yaitu menormalisasi block menggunakan L2, lalu membatasi maksimum value menjadi 0.2, lalu dinormalisasi kembali menggunakan L2. L2 normalization bekerja dengan cara membagi value dengan hasil akar kuadrat dari jumlah kuadrat tiap value.

$$\text{L2 norm} = v \leftarrow v / \sqrt{\|v\|_2^2 + \epsilon^2}$$

```
hog_feature = hog(image, orientations=9, pixels_per_cell=(4, 4),
cells_per_block=(8, 8), visualize=False)
```

Nilai parameter pixels_per_cell (ppc) dan cells_per_block (cpb) digunakan karena memiliki akurasi tertinggi saat percobaan menggunakan 2199 testing data.

- **Classifier**

Classifier yang digunakan adalah Linear Support Vector Classification, karena keterbatasan hardware yang mengurungkan saya mencoba classifier lain seperti Stochastic Gradient Descent Classifier atau Neural Networks.

- **Testing & Akurasi**

Menggunakan 2199 testing data (200 data tiap digit), akurasi tertinggi diperoleh dari ppc=(4, 4), cpb=(8, 8).

Pixels_per_cell	Cells_per_block	Accuracy
(2, 2)	(8, 8)	0.9358799454297408
(2, 2)	(16, 16)	0.9540700318326512
(4, 4)	(4, 4)	0.9586175534333788
(4, 4)	(8, 8)	0.9618008185538881
(8, 8)	(4, 4)	0.9477035015916325

Testing Dunia Nyata

Walaupun classifier mendapat skor tinggi saat testing menggunakan 2199 testing data, nyatanya hanya berkisar 75-90%. Hal ini diakibatkan proses ekstrak angka dari form C1 yang kaku.



Kesalahan potong mengakibatkan classifier gagal memprediksi data.

Saran

1. Diperlukan digit extractor yang lebih baik, mayoritas terjadi misprediksi dikarenakan salah potong oleh digit extractor kami.
2. Mode Otomatis sehingga user tidak perlu menunjukkan aplikasi halaman TPS secara manual. (Rencana awal VSP-2019, namun karena masalah stabilitas, bentuk aplikasi VSP-2019 diubah menjadi versi saat ini).
3. Menambah kelas tambahan, yaitu “-“. Karena banyak petugas KPPS menulis X dengan “-“.
4. Menambah Exception “0” = “X”. Karena banyak petugas KPPS menulis X X X pada nilai 0, semestinya X X O.
5. Berhubungan dengan poin 3 dan 4, mengedukasi para petugas KPPS pentingnya menulis hasil perhitungan suara dengan teratur, bukan hanya aplikasi sulit memprediksi, manusia juga.

Source Code

<https://github.com/ouwyukha/ValidatorPilpres2019-dengan-HOG-LinearSVC>

RAW Dataset

https://drive.google.com/file/d/1HY1pl8rn60tfArClkP9A9eM_yab3SKRe/view?usp=sharing

Referensi

<https://pemilu2019.kpu.go.id/#/ppwp/hitung-suara/>

<https://selenium-python.readthedocs.io/api.html>

<https://blog.chromium.org/2011/05/remote-debugging-with-chrome-developer.html>

<https://www.programcreek.com/python/example/89830/win32gui.DestroyWindow>

<https://www.learnopencv.com/image-alignment-feature-based-using-opencv-c-python/>

<https://stackoverflow.com/questions/50759678/cropping-greyscale-images-within-larger-images-using-opencv-and-python/50760599>

<https://github.com/bikz05/digit-recognition>

<https://www.learnopencv.com/histogram-of-oriented-gradients/>

<https://tel.archives-ouvertes.fr/tel-00390303/file/NavneetDalalThesis.pdf>

https://scikit-image.org/docs/dev/auto_examples/features_detection/plot_hog.html

<https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearSVC>