

# Metode Pengumpulan Data Pada Deteksi Pakaian Hijab Syar'I Berdasarkan Citra Digital Menggunakan *Teachable machine* Learning

<sup>1</sup>Fathorazi Nur Fajri, <sup>2</sup>Kamil Malik, <sup>3</sup>Gulpi Qorik Oktagalu Pratamasunu

<sup>1</sup>Sistem Informasi, Universitas Nurul Jadid, Indonesia

<sup>2,3</sup>Informatika, Universitas Nurul Jadid, Indonesia

[fathorazi@unuja.ac.id](mailto:fathorazi@unuja.ac.id), [kamil@unuja.ac.id](mailto:kamil@unuja.ac.id), [pratamasunu@unuja.ac.id](mailto:pratamasunu@unuja.ac.id)

## ARTICLE INFO

### Article History:

Diterima : 25-10-2022

Disetujui : 04-11-2022

### Keywords:

Hijab Syar'I;

Citra digital;

Machine Learning;

Teachable Machine;

Computer vision



## ABSTRACT

**Abstract:** Computer vision is a branch of artificial intelligence that uses digital images as input data. The stages of making computer vision are data collection, preprocessing, modeling, and testing. Test results are usually influenced by good modeling and data collection. Errors often occur in the data collection process so that data collection testing is needed before the modeling process using matlab or python. Therefore, this research aims to propose a data collection model with testing using a teachable machine. The datasets used are hijab syar'i, non syar'i and non hijab datasets because hijab is a trend among young people. However, many young people are not good at using their hijab in the pesantren area. So that the classification process is needed as a form of order in the pesantren. Datasets are obtained using web scraping methods on internet platforms such as pinterest, google image and instagram. The total dataset is 2,910 digital images of each class is 970. Teachable machine can validate the dataset quickly by looking at the accuracy per class. With the best model in the configuration of epoch 50, batch size 64 and learning rate 0.0001. Where the accuracy of each class using teachable is hijab syar'i by 99%, non syar'i by 94% and non hijab 99% with a test sample of 146 digital images each.

**Abstrak:** Visi Komputer merupakan cabang kecerdasan buatan yang menggunakan citra digital sebagai input data. Adapun tahapan pembuatan visi komputer ialah pengumpulan data, preprocessing, pembuatan model, dan pengujian. Hasil pengujian biasanya dipengaruhi oleh pembuatan model dan pengumpulan data yang baik. Kesalahan sering terjadi pada proses pengumpulan data sehingga dibutuhkan pengujian pengumpulan data sebelum proses pembuatan model dengan menggunakan matlab atau python. Oleh karenanya pada penelitian ini bertujuan untuk mengusulkan model pengumpulan data dengan pengujian menggunakan teachable machine. Adapun dataset yang digunakan ialah dataset hijab syar'i, non syar'i dan non hijab dikarenakan hijab menjadi trend dikalangan anak muda. Akan tetapi banyak anak muda yang belum baik penggunaan hijabnya diarea pesantren. Sehingga diperlukan proses klasifikasi sebagai bentuk ketertiban didalam pesantren. Dataset didapat dengan menggunakan metode web scraping pada platform internet seperti pinterest, google image dan instagram. Total dataset ialah 2,910 citra digital masing masing class ialah 970. Teachable machine dapat melakukan validasi dataset secara cepat dengan melihat akurasi per class. Dengan model terbaik pada konfigurasi epoch 50, batch size 64 dan learning rate 0.0001. Dimana Akurasi setiap class menggunakan teachable yaitu hijab syar'i sebesar 99%, non syar'i sebesar 94% dan non hijab 99% dengan sampel uji masing masing 146 citra digital.



<https://doi.org/10.31764/justek.vXIY.ZZZ>



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

## A. LATAR BELAKANG

Islam begitu memuliakan dan menyanjung kaum wanita. Wanita disandingkan dengan wewangian dan perhiasan yang identik dengan keindahan, keanggunan dan kemewahan. Mengenai kecantikan Islam mewajibkan kaum hawa untuk menutup aurat. Aurat merupakan hal yang dapat menimbulkan birahi/syahwat sehingga harus ditutup dan dipelihara agar tidak mengganggu dan menimbulkan kemurkaan. Aurat bagi muslimah adalah semua bagian tubuh kecuali wajah dan telapak tangan (Karlina, 2018). Hijab syar'i merupakan cara berpakaian bagi seorang wanita muslim untuk menutup auratnya dimana tidak terlihatnya lekuk tubuh. Hijab syar'i bukanlah hijab atau jilbab yang modis atau tren akan tetapi jilbab besar yang dapat menutup aurat selain itu tidak terawang dan tidak juga memperlihatkan lekuk tubuh (Daud, 2013). Banyaknya tren hijab saat ini terkadang membuat masyarakat sangat sulit untuk membedakan antara pakaian syar'i dan non-syar'i, apalagi di era saat ini banyak *style* jilbab modis yang populer di kalangan anak muda.

Saat ini, perkembangan teknologi komputer semakin berkembang menyebabkan citra digital menjadi lebih banyak dan kompleks. Citra digital saat ini digunakan di berbagai bidang seperti keamanan, perdagangan, perkebunan (Fajri, Hamid, & Pramunendar, 2017) dan banyak lagi penggunaan informasi dalam bentuk gambar di bidang lain. Terdapat banyak platform yang menyediakan citra digital seperti pinterest, google image, instagram dan media sosial lainnya. Pemrosesan gambar dapat dilakukan dengan meningkatkan kualitas gambar yang ada selain itu terdapat ilmu-ilmu komputasi yang memungkinkan manusia untuk mendapatkan sebuah informasi dari sebuah citra digital atau gambar (Gonzalez & Woods, 2018) Bidang Machine learning sendiri terus berkembang agar komputer dapat meniru cara kerja indera manusia khususnya deteksi objek.

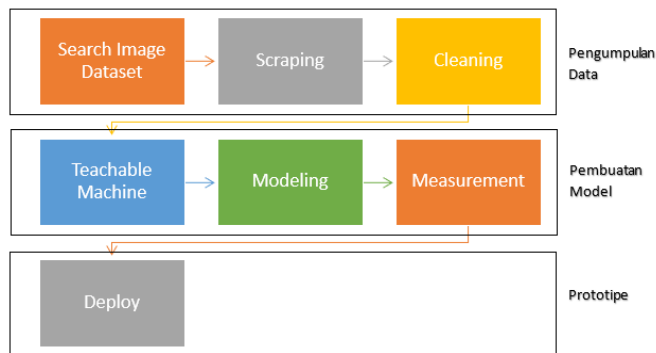
Beberapa tahun terakhir ini topik hijab syar'i menarik untuk dilakukan sebuah penelitian. Seperti penelitian terkait mix and match hijab berbasis android (Susanti, 2015), pemilihan model hijab (Sari, Permana, & Sali, 2017), dan deteksi wajib berhijab menggunakan deep learning (Anggraini, 2020). Setiap hari banyak muslimah yang berkunjung ke pondok pesantren nurul jadid sehingga terkadang membuat kesulitan untuk mendeteksi hijab yang syar'i para pengunjung.

Membuat sebuah sistem computer vision dibutuhkan proses yang panjang terlebih pada proses pengumpulan dataset yang mana memungkinkan dataset tersebut tidak valid dan hal tersebut baru disadari ketika proses pembuatan model computer vision. Sehingga dibutuhkan sebuah metode yang dapat mendeteksi secara dini validitas dataset tersebut dengan cepat dan mudah. Oleh karenanya pada penelitian ini mengusulkan tentang proses pengumpulan dataset dan pengguna Teachable machine untuk menguji validitas dataset pada deteksi hijab syar'i. Sehingga proses kesalahan pengumpulan data yang dapat mempengaruhi akurasi pada proses pembuatan model dapat diketahui lebih awal.

## B. METODE PENELITIAN

Pada penelitian terdapat 2 tahapan yaitu metode pengumpulan data dan metode pengolahan data sehingga menghasilkan model dan *prototipe*. Adapun aktivitas pada penelitian tersebut ialah sebagai berikut:

1. Mencari dataset citra digital pada penyedia *platform* seperti *pinterest*, *instagram*, *google image* dan *platform* lainnya.
2. Melakukan *scraping* untuk mengunduh hasil pencarian pada *platform*
3. Melakukan proses *cleaning* dataset yaitu menghapus gambar *duplicate*, mengelompokkan dataset berdasarkan kategori atau kelasnya.
4. Mengupload dataset pada *Teachable machine*
5. Melakukan *Tune up modeling* seperti *batch size*, *epoch* dan *learning rate* untuk mendapatkan model dengan akurasi terbaik.
6. Melakukan pengukuran pada model yang dihasilkan dengan melihat akurasi klasifikasi gambar yang dihasilkan.
7. Melakukan proses *deploy* untuk mendapatkan protipe model.



**Gambar 1.** Metode Penelitian

Sebagai bahan literatur pada penelitian ini dibahas juga terkait hijab syar’I, *scraping*, *Teachable machine* dan metode pengukuran.

### 1. Hijab Syar’i

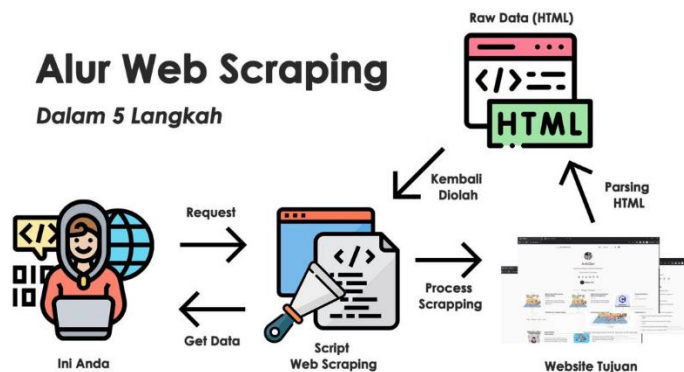
Hijab syar’i merupakan hijab yang di syaratkan tidak boleh tipis, berbentuk lekuk tubuh karena dapat mendatangkan fitnah baik bagi sendiri maupun bagi orang lain. Adapun perbedaan hijab biasa dengan hijab syar’i seperti pada gambar.



**Gambar 2.** Perbedaan Hijab Syar’I dengan Hijab biasa

## 2. Scraping

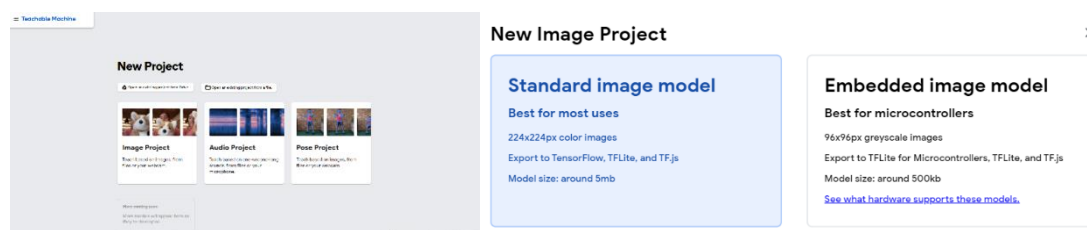
Proses pengambilan informasi dari website-website yang ada ini disebut dengan “web scraping”. *Web scraping* adalah teknik untuk mendapatkan informasi dari website secara otomatis tanpa harus menyalinnya secara manual. Cara kerja web scraping dalam mendapatkan data dengan cara pengambilan dan ekstraksi. Saat ini *web scraping* telah digunakan dalam pengumpulan data cepat pada *marketplace*, artikel dan lain lain (AYani, Pratiwi, & Muhardi, 2019) (Josi, Abdillah, & Suryayusra, 2014).



Gambar 3. Cara Kerja Scraping

## 3. Teachable Machine

*Teachable machine* adalah antarmuka berbasis web yang memungkinkan orang untuk melatih model klasifikasi *Machine learning* mereka sendiri tanpa coding dengan menggunakan *webcam*, gambar, atau suara (Carney, Webster, Alvarado, Phillips, & Howell, 2020). *Teachable machine* menggunakan *transfer learning* sebagai teknik *machine learning* untuk menemukan pola dan *trend* dalam sample gambar atau suara sehingga membuat model klasifikasi dengan mudah dan cepat. Dengan *transfer learning* memungkinkan pengguna menambahkan sendiri dataset dan melatih kembali model sebelumnya. *Teachable machine* dirancang agar dapat digunakan para pendidik untuk dengan mudah mengajarkan konsep *machine learning* seperti deployment model, test cepat data penelitian dan memungkinkan pengguna disabilitas untuk membuat model *machine learning* sendiri (Harditya, 2020).



Gambar 4. Tampilan Web Base Teachable Machine

## 4. Measurement

Metode *Measurement* atau pengukuran yang digunakan dalam penelitian ini adalah recall, precision, dan accuracy (Novianti, Pradnyani, & Setiawa, 2015). Recall dan

precision merupakan sebuah rasio prediksi yang dilakukan terhadap kelas positif, yang berfungsi sebagai pengukur seberapa tepat dan lengkap klasifikasi yang telah dilakukan. Berikut ini adalah acuan *confusion matrix* yang memperlihatkan prediksi dan kondisi sebenarnya dari data yang dihasilkan oleh algoritma yang digunakan.

**Tabel 1.** *Confusion matrix* dari suatu klasifikasi

Predicted value	Actual Value	
	Positif	Negative
Positif	TP	FP
Negative	FN	TN

Berdasarkan *confusion matrix* diatas yaitu *recall* (r) dan *precision* (p) kelas positif dapat dihitung dengan menggunakan rumus. *recall* (r) merupakan jumlah prediksi yang benar dalam kelas positif dibagi dengan jumlah positif yang sebenarnya dalam penelitian. Presisi (p) adalah jumlah prediksi yang benar di kelas yang positif dibagi dengan jumlah diklasifikasi sebagai positif.

$$precision = \frac{\sum_{i=1}^i TP_i}{\sum_{i=1}^i (FP_i + TP_i)} \times 100\% \quad (1)$$

$$Recall = \frac{\sum_{i=1}^i TP_i}{\sum_{i=1}^i (TP_i + FN_i)} \times 100\% \quad (2)$$

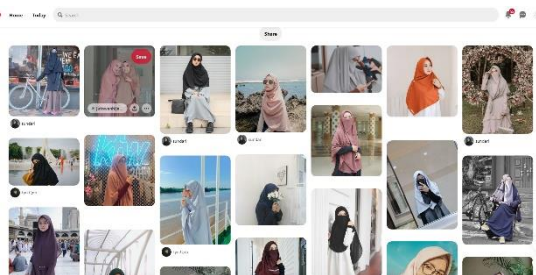
*Accuracy* adalah rasio prediksi yang benar (positif dan negatif) dengan data keseluruhan. Sehingga akurasi ini dapat menjawab prediksi yang benar dari keseluruhan data yang ada. Rumus yang digunakan dalam *accuracy* adalah sebagai berikut:

$$Accuracy = \frac{\sum_{i=1}^i \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{1} \times 100\% \quad (3)$$

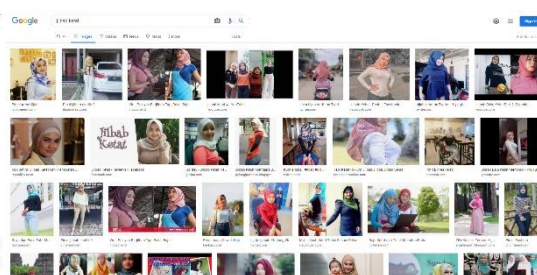
## C. HASIL DAN PEMBAHASAN

### 1. Pengumpulan Dataset

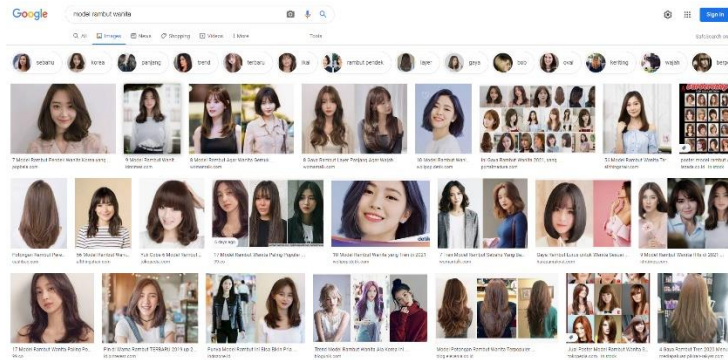
Pengumpulan dataset citra digital hijab syar'I, non syar'I dan non hijab bersumber pada media social dan *search engine*. Dataset hijab syar'I diperoleh pada *platform pinterest* dengan *keyword* "hijab syar'I". Dataset hijab non syar'I diperoleh pada *platform google image* dengan *keyword* "jilbab modis", "jilboob", dan "jilbab ketat". Kemudian untuk dataset non hijab diperoleh pada platform google image dengan *keyword* model rambut wanita.



**Gambar 5.** Pinterest Hijab Syar'I



**Gambar 6.** Google Image Jilbab Ketat



**Gambar 7.** Google Image Model Rambut Wanita

Setelah mendapatkan *keyword* yang tepat pada *search engine pinterest* dan *google image* dilakukan proses *scraping* untuk mendapatkan citra digital tersebut sehingga mengurangi proses manual dan menghemat waktu pengumpulan dataset. Untuk mendapatkan *dataset* yang baik maka dilakukan proses *cleaning* yaitu dengan cara

1. Menghapus gambar yang sama
2. Menghapus gambar yang mempunyai teks terlalu banyak
3. Mengelompokkan gambar sesuai kategorinya yaitu hijab syar'I, non syar'I dan non hijab
4. Menghapus gambar yang tidak termasuk pada kategori atau class

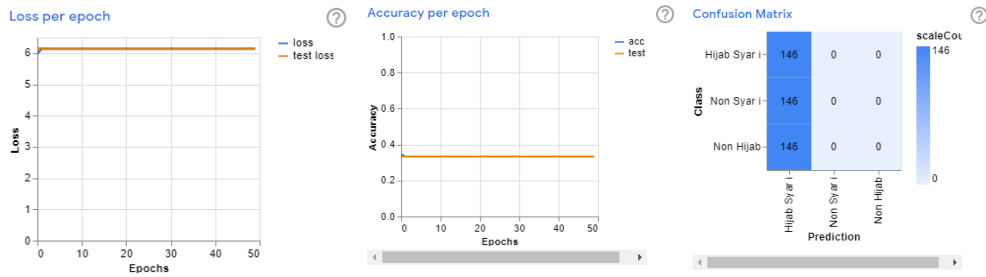
Sehingga diperoleh dataset citra digital sebanyak 970 gambar pada masing masing kategori hijab syar'I, non syar'I dan non hijab. Berikut adalah hasil citra digital yang telah dilakukan proses *cleaning*. Komposisi dataset *training* dan *testing* ialah 85% data *training* dan 15% data *testing*.

## 2. Pembuatan Model

Pembuatan model dilakukan dengan menggunakan *tool web base machine learning* yaitu *teachable machine*. Proses pertama ialah mengupload dataset yang telah dikumpulkan sesuai dengan *class* nya. Selanjutnya melakukan proses konfigurasi model latih seperti *epoch*, *batch size* dan *learning rate*. Secara default konfigurasi model *Teachable machine* yaitu epoch 50, batch size 16 dan learning rate 0.001.

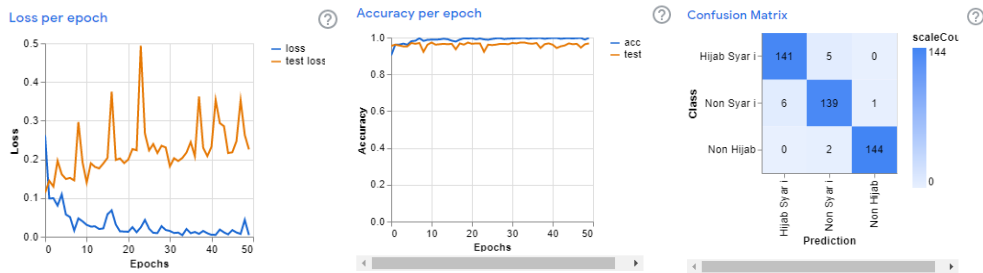
Pada penelitian ini terdapat 3 class atau kategori yaitu hijab syar'I, non syar'I dan non hijab. Untuk mendapatkan model terbaik maka dilakukan grid search sehingga bisa terhindar dari *underfitting* ataupun *overfitting*. Proses *grid search* dilakukan dengan *tune up* pada parameter tersebut. Adapun konfigurasi *tune up* yang dilakukan ialah *epoch* dari 10 – 50, *batch size* dari 16 - 64 dan *learning rate* 0.1 – 0.0001. Percobaan pertama *tune up* pada epoch 50, *batch size* 16 dan *learning rate* 0.1 model belum bisa menganalisa dataset. Dikarenakan *loss* sangat tinggi dan akurasi yang rendah.





Gambar 8. Model Percobaan pertama

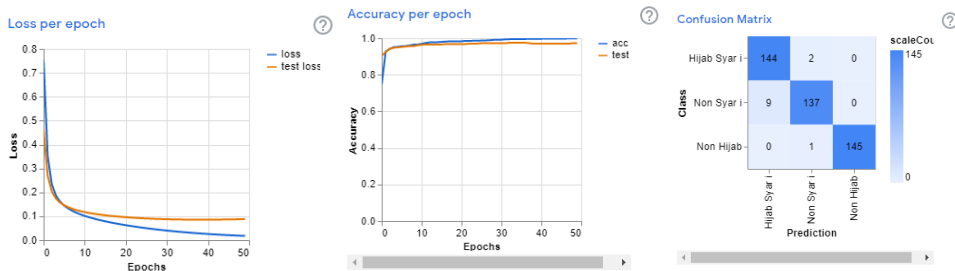
Pada percobaan selanjutnya dilakukan konfigurasi dengan menurunkan *learning rate* menjadi 0.01 dengan *epoch* 50 dan *batch size* 16. Menghasilkan model yang *overfitting* dimana *loss error* pada *training* dataset rendah dan *error test dataset* tinggi. *Overfitting* juga terjadi ketika menaikkan nilai *batch size* menjadi 32 dan *learning learning* 0.001. Percobaan terus dilakukan sampai mendapatkan model terbaik pada konfigurasi *epoch* 50, *batch size* 64 dan *learning learning* 0.0001.



Gambar 9. Model Overfitting

### 3. Pengukuran Evaluasi Model

Pada proses pembuatan model didapatkan model terbaik pada konfigurasi *epoch* 50, *batch size* 64 dan *learning rate* 0.0001. Dimana *loss* pada data *training* rendah dan *loss* pada data *testing* rendah terlihat pada grafik yang menunjukkan terjadi pembelajaran pada model. Akurasi yang dihasilkan pada model 97% dan nilai *loss* 0.08.



Gambar 10. Model terbaik Hijab Syar'i

Accuracy per class ?

CLASS	ACCURACY	# SAMPLES
Hijab Syar i	0.99	146
Non Syar i	0.94	146
Non Hijab	0.99	146

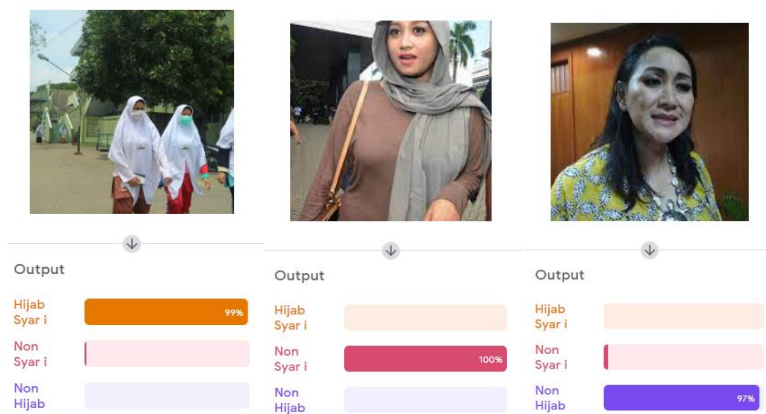
**Gambar 11.** Akurasi Pada Setiap Kategori

#### 4. Deploy / Prototipe

Proses *Deploy* atau pembuatan *prototipe* dapat dilakukan dengan memanfaatkan fitur *export* model pada *teachable machine*. Dimana pada banyak opsi *export* model seperti proses *export* model ke *tensorflow*, *ml5js*, *p5.js*, *coral*, *framer*, *node*, *android tensorflow lite*. Pada penelitian ini fitur *export* model yang digunakan ialah *export model cloud shareable* yang mana secara otomatis model yang telah kita buat dibuatkan url protipe yang langsung bisa digunakan. Inputan pada protipe bisa berupa webcam dan upload gambar. Dengan tampilan design yang *hybrid* atau dinamis maka *prototipe* juga bisa langsung digunakan pada *smartphone* dengan menggunakan kamera *smartphone*.

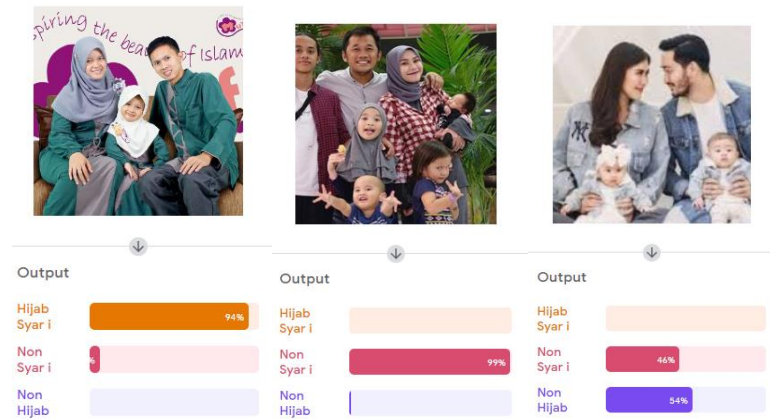
#### D. TEMUAN ATAU DISKUSI

Hasil ujicoba dilakukan dengan cara upload citra digital atau menggunakan kamera. percobaan dilakukan dengan citra digital sederhana atau perorang dan citra digital kompleks atau foto keluarga.



**Gambar 12.** Hasil akurasi model pada sample sederhana





**Gambar 13.** Hasil akurasi model pada sample kompleks

Terdapat limitasi atau keterbatasan pada model ini sehingga menyebabkan ketidakakuratan deteksi hijab syar'i. limitasi pada model ini ialah terletak pada jarak atau jauh dekatnya pengambilan citra digital untuk non hijab. Sehingga jika citra digital *fullbody* maka model kesulitan untuk mendeteksi non hijab.



**Gambar 14.** Tidak Akurat Citra Fullbody



**Gambar 15.** Akurat Citra Fokus wajah

## E. SIMPULAN DAN SARAN

Kesimpulan pada penelitian ini ialah teachable machine dapat digunakan sebagai tool untuk melakukan validasi dataset dengan melihat akurasi pada setiap class dataset. Dengan begitu validasi dataset dapat dilakukan secara cepat. Adapun model transfer learning terbaik dari teachable machine ialah Dengan konfigurasi epoch 50, batch size 64 dan learning rate 0.0001. Dimana Akurasi setiap class menggunakan teachable yaitu hijab syar'i sebesar 99%, non syar'i sebesar 94% dan non hijab 99% dengan sampel uji masing masing 146 citra digital. Selain itu Fitur export cloud memungkinkan untuk dapat secara langsung melakukan proses deploy.

Adapun saran untuk penelitian selanjutnya ialah dapat menggunakan dataset berikut dikarenakan dataset berikut sudah cukup baik untuk mendeteksi model hijab syar'i hanya saja terdapat limitasi jika input gambar testing menggunakan gambar

fullbody maka model belum dapat menghasilkan deteksi yang akurat. Sehingga sebelum mendeteksi model hijab dapat dilakukan deteksi wajah terlebih dahulu.

## REFERENSI

- Anggraini, W. (2020). Deep Learning Untuk Deteksi Wajah Yang Berhijab Menggunakan Algoritma Convolutional Neural Network (CNN) Dengan Tensorflow. Banda Aceh: UIN Ar-Raniry.
- AYani, D. D., Pratiwi, H. S., & Muhandi, H. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. JUSTIN (Jurnal Sistem dan Teknologi Informasi), 7(4), 257-262.
- Carney, M., Webster, B., Alvarado, I., Phillips, K., & Howell, N. (2020). Teachable machine: Approachable Web-based tool for exploring machine learning classification. Extended abstracts of the 2020 CHI conference on human factors in computing systems.
- Daud, F. K. (2013). Jilbab, hijab dan aurat perempuan (Antara tafsir klasik, tafsir kontemporer dan pandangan muslim feminis). Al Hikmah: Jurnal Studi Keislaman, 3(1), 1-1.
- Fajri, F. N., Hamid, N., & Pramunendar, R. A. (2017). The recognition of mango varieties based on the leaves shape and texture using back propagation neural network method. In 2017 International Conference on Sustainable Information Engineering and Technology (SIET). Malang.
- Gonzalez, R. C., & Woods, R. E. (2018). Digital Image Processing, Hoboken. Pearson.
- Harditya, A. (2020). Indonesian Sign Language (BISINDO) As Means to Visualize Basic Graphic Shapes Using Teachable Machine. International Conference of Innovation in Media and Visual Design (IMDES 2020) Atlantis Press.
- Josi, A., Abdillah, L. A., & Suryayusra. (2014). Penerapan teknik web scraping pada mesin pencari artikel ilmiah. arXiv preprint arXiv:1410.5777.
- Karlina, L. (2018). Pemakaian hijab syar'i mahasiswi IAIN Palangka Raya. (Doctoral dissertation, IAIN Palangka Raya).
- Novianti, Pradnyani, K. D., & Setiawa, N. A. (2015). Peningkatan Nilai Recall dan Precision pada Penelusuran Informasi Pustaka Berbasis Semantik (Studi Kasus: Sistem Informasi Ruang Referensi Jurusan Teknik Elektro dan Teknologi Informasi UGM). Proceedings Konferensi Nasional Sistem dan Informatika (KNS&I).
- Sari, U. I., Permana, I., & Sali, F. N. (2017). Sistem Pendukung Keputusan Berbasis Rule Untuk Pemilihan Model Hijab. Seminar Nasional Teknologi Informasi Komunikasi Dan Industri.
- Susanti, S. (2015). Rancang Bangun Aplikasi Mix dan Match Hijab dengan Android face detector Api. Makassar: Universitas Islam Negari Alauddin.