

PENGEMBANGAN PROYEK *MACHINE LEARNING* PREDIKSI TUBERKULOSIS PADA KEGIATAN MBKM STUDI INDEPENDEN BERSERTIFIKAT DICODING 2021

M. Audina¹, A.A.I.N.E. Karyawati², dan I.W. Supriana³

ABSTRAK

Kampus Merdeka merupakan bagian dari kebijakan Merdeka Belajar oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia yang memberikan kesempatan bagi mahasiswa untuk mengasah kemampuan sesuai minat dan bakat dengan terjun langsung ke dunia kerja sebagai persiapan karier di masa depan. Salah satu program yang ditawarkan adalah Merdeka Belajar Kampus Merdeka Studi Independen Bersertifikat (MBKM SIB) pada PT. Presentologics Dicoding Indonesia. Satu dari empat paket belajar yang disediakan adalah paket Pengembang *Machine Learning* dan *Front-End Web*. *Machine learning* terus berkembang setiap tahunnya dan semakin banyak diimplementasikan di berbagai bidang industri, salah satunya adalah bidang kesehatan. Pada proyek akhir dari kegiatan MBKM SIB Dicoding ini, telah dikembangkan proyek *machine learning* yaitu proyek prediksi penyakit tuberkulosis berdasarkan data gambar *chest x-ray* yang dapat mengklasifikasikan dan memprediksikan data *chest x-ray* ke dalam dua kategori yaitu normal dan tuberkulosis. Proyek ini diharapkan dapat membantu dokter atau tenaga ahli kesehatan sebagai pendukung keputusan dalam proses diagnosis penyakit tuberkulosis melalui citra *chest x-ray* pasien.

Kata kunci : Tuberkulosis, Prediksi, Klasifikasi, Machine Learning, Chest X-Ray

ABSTRACT

Kampus Merdeka is part of the Merdeka Belajar policy by the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia which provides opportunities for students to hone skills according to their interests and talents by going directly into the world of work as preparation for future careers. One of the programs offered is Merdeka Belajar Kampus Merdeka Studi Independen Bersertifikat (MBKM SIB) at PT. Presentologics Dicoding Indonesia. One of the four learning packages provided is Machine Learning and Front-End Web Developer package. Machine learning continues to grow every year and is increasingly being implemented in various industrial fields, one of which is the healthcare sector. In the final project of the MBKM SIB Dicoding activities, a machine learning project has been developed, namely the prediction of tuberculosis based on chest x-ray image data that can classify and predict chest x-ray data into two categories, namely normal and tuberculosis. This project is expected to help doctors or health professionals as decision support in the process of diagnosing tuberculosis through chest x-ray images of patients.

Keywords : Tuberculosis, Prediction, Classification, Machine Learning, Chest X-Ray

1. PENDAHULUAN

Merdeka Belajar Kampus Merdeka Studi Independen Bersertifikat (MBKM SIB) adalah program yang diselenggarakan oleh Kementerian Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik

¹ Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana, Jl. Raya Kampus Unud, Jimbaran, 80361, Badung-Bali, marissaaudina@gmail.com

² Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana, Jl. Raya Kampus Unud, Jimbaran, 80361, Badung-Bali, eka.karyawati@unud.ac.id

³ Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana, Jl. Raya Kampus Unud, Jimbaran, 80361, Badung-Bali, wayan.supriana@unud.ac.id

Submitted: 7 November 2022

Revised: 25 November 2022

Accepted: 27 November 2022

Indonesia. Salah satu mitra pada kegiatan ini adalah PT. Presentologics Dicoding Indonesia yang menyediakan empat paket belajar bagi mahasiswa peserta MBKM SIB. Satu dari keempat paket belajar yang ditawarkan adalah paket Pengembangan *Machine Learning* dan *Front-End Web*. Tantangan bagi para *developer* yang berkecimpung langsung di dunia teknologi khususnya *machine learning* adalah materi-materi pembelajaran yang memiliki standar tinggi sesuai dengan standar industri. Oleh karena itu, Dicoding bersama perusahaan teknologi bekerja sama untuk menghadirkan materi pembelajaran berkualitas tinggi tersebut melalui *platform* Dicoding Academy (Dicoding, 2021).

Pada program ini, mahasiswa dapat menambah pengalaman baik secara akademik maupun non-akademik. Mahasiswa mendapatkan fasilitas dan akses pada kelas dan modul-modul yang terdapat di *platform* Dicoding. Terdapat total 14 kelas selama program ini berlangsung yang mencakup materi dasar pemrograman, *machine learning*, dan *front-end web*. Selain kelas dan modul pembelajaran materi, kegiatan tatap muka secara daring juga dilakukan untuk menambah pemahaman materi yang dibawakan langsung oleh para ahli di bidangnya. Secara non-akademik, mahasiswa juga diajarkan dan dibimbing untuk siap menghadapi dunia kerja melalui materi-materi *soft skill* seperti *personal branding*, *critical thinking*, dan lain-lain.

Sesuai dengan paket belajar, *machine learning* banyak dibahas pada kelas dan modul yang disediakan. *Machine learning* merupakan cabang dari *Artificial Intelligence* (AI) dan ilmu komputer yang berfokus pada penggunaan data dan algoritma untuk meniru cara manusia belajar, kemudian secara bertahap meningkatkan akurasi (IBM Cloud Education, 2020). Perkembangan pesat dari waktu ke waktu membuat *Machine learning* semakin banyak diimplementasikan di berbagai bidang industri, salah satunya adalah bidang kesehatan. *Machine learning* telah banyak dimanfaatkan dalam bidang kesehatan khususnya pada klasifikasi atau deteksi penyakit, seperti klasifikasi tumor otak pada citra MRI (Febrianti, Sardjono, & Babgei, 2020), deteksi COVID-19 menggunakan gambar *chest x-ray* (Nayak et al, 2021), lain-lain.

Pada proyek akhir program ini, telah dibangun dan dikembangkan suatu proyek *machine learning* yang dapat memprediksikan penyakit tuberkulosis berdasarkan gambar *chest x-ray*. Proyek ini dilatar belakangi oleh fakta bahwa penyakit tuberkulosis merupakan salah satu penyakit penyebab kematian tertinggi di Indonesia. Menurut Data Global TB *record* 2021, jumlah kasus tuberkulosis yang ditemukan dan diobati di Indonesia mencapai 393.323 kasus di tahun 2020 dan 209.575 kasus per 4 Oktober 2021 (Kemenkes, 2021). Sekitar seperempat populasi dunia terinfeksi tuberkulosis (World Health Organization, 2021). Tuberkulosis atau biasa disebut dengan TBC adalah penyakit menular yang disebabkan oleh adanya infeksi bakteri atau kuman *Mycobacterium Tuberculosis* (Chomaerah, 2020). Salah satu cara untuk mendeteksi penyakit tuberkulosis adalah dengan melihat hasil rontgen toraks atau *chest x-ray*. Namun, membaca *chest x-ray* adalah tugas yang sulit, menantang, dan masih dianggap penting meskipun telah banyak perkembangan teknik pencitraan yang kuat seperti *computed tomography*, *computed tomography* beresolusi tinggi, dan resonansi magnetik (Delrue et al, 2011). Pada proyek ini, model *machine learning* yang dibangun mengimplementasikan algoritma *Convolutional Neural Network* yang dapat mengklasifikasikan dan memprediksikan data *chest x-ray* ke dalam dua kategori yaitu normal dan tuberkulosis. Proyek ini diharapkan dapat membantu dokter atau tenaga ahli kesehatan dalam proses diagnosis penyakit tuberkulosis melalui citra *chest x-ray* pasien.

2. METODE PELAKSANAAN

2.1 Pelaksanaan Kegiatan

Program Studi Independen Dicoding pada paket belajar Pengembangan *Machine Learning* dan *Front-End Web* ini berlangsung dari 23 Agustus 2021 – 19 Januari 2022. Terdapat beberapa metode

pembelajaran yang dilakukan seperti *self learning* dengan mengakses dan menyelesaikan total 14 kelas serta modul-modul *machine learning* dan *front-end web* pada *platform* Dicoding. Selain itu, dilakukan juga pembelajaran tatap muka secara daring melalui Google Meet yang dilaksanakan selama dua minggu sekali. Setelah seluruh kelas selesai dan mahasiswa mendapatkan pengetahuan serta keahlian di bidang yang terkait, *output* akhir dari program ini adalah membuat suatu proyek *machine learning* atau *front-end web* dengan tema-tema tertentu yang sudah ditentukan. Tema yang dipilih pada proyek ini adalah tema kesehatan diri dan mental. Proyek akhir yang telah dibangun ini berusaha untuk mengatasi permasalahan di bidang kesehatan yaitu kesulitan dalam interpretasi *chest x-ray* diagnosis tuberkulosis yang dilakukan oleh dokter atau ahli radiologis. Sehingga, dengan adanya proyek ini diharapkan dapat menjadi pendukung keputusan dokter dalam mendiagnosis penyakit tuberkulosis.

2.2 Struktur Penyusunan Proyek *Machine Learning*

Berdasarkan permasalahan bidang kesehatan yang dihadapi, maka solusi pemecahan masalah tersebut yaitu dengan membuat sebuah proyek *machine learning* dengan beberapa tahapan yaitu sebagai berikut : 1) Mendefinisikan *problem statement*. 2) Melakukan eksplorasi dan persiapan data. 3) Membangun model *machine learning*. 4) Evaluasi model yang sudah dibangun. 5) Melakukan *deployment* sebagai tahap finalisasi. 6) Pengujian sistem yang sudah dibangun.

3. HASIL DAN PEMBAHASAN

Kegiatan MBKM Studi Independen Bersertifikat Dicoding ini sudah berjalan dengan baik dan lancar. Seluruh kelas telah terselesaikan termasuk tugas-tugas dan proyek-proyek yang diberikan. Selain itu, seluruh sesi tatap muka secara daring juga sudah dilakukan dan berhasil menambah ilmu pengetahuan serta pengalaman untuk mahasiswa. Sebagai *output* akhir dari program ini, telah dibangun sebuah proyek *machine learning* yang dapat mengklasifikasikan dan memprediksikan penyakit tuberkulosis berdasarkan gambar *chest x-ray*. Berikut adalah tahapan dalam menyusun proyek *machine learning* ini.

3.1 *Problem Statement*

Tuberkulosis merupakan salah satu penyakit penyebab kematian tertinggi di Indonesia. Salah satu cara untuk mendeteksi penyakit tuberkulosis adalah dengan melihat hasil rontgen toraks atau *chest x-ray*. Membaca dan menginterpretasikan *chest x-ray* adalah tugas yang tidak mudah bagi seorang dokter, ahli radiologis, atau tenaga ahli kesehatan lainnya. Oleh karena itu, proyek ini dibuat dengan memanfaatkan *machine learning* dalam membangun model yang dapat mengklasifikasikan dan memprediksikan gambar *chest x-ray* tersebut. Proyek ini diharapkan dapat dimanfaatkan dan berguna sebagai pendukung keputusan dokter atau tenaga ahli kesehatan dalam proses diagnosis tuberkulosis.

3.2 Eksplorasi dan Persiapan Data

Tahap eksplorasi data dilakukan untuk memahami bagaimana sebaran data, jenis data, dan sebagainya. Pada proyek ini, data yang digunakan merupakan data sekunder yang bersumber dari penelitian tuberkulosis yang dilakukan oleh (Rahman *et al*, 2020) yang dapat diakses melalui situs Kaggle pada link <https://www.kaggle.com/tawsifurrahman/tuberculosis-tb-chest-xray-dataset>. Dataset yang digunakan merupakan data gambar *chest x-ray* yang terdiri dari 3500 data normal dan 700 data pengidap tuberkulosis.

Tahap selanjutnya adalah persiapan data. Sebelum data masuk pada proses pemodelan, dilakukan proses *undersampling* terlebih dahulu, yaitu proses yang dilakukan untuk menggunakan data sedemikian rupa sehingga proporsi kelas mayoritas yaitu kelas normal tidak sebesar proporsi aslinya dan menjadi seimbang dengan kelas minoritas yaitu kelas tuberkulosis. Kemudian, data gambar *chest*

x-ray diproses kembali dalam tahapan *image augmentation* yaitu teknik mengubah data gambar yang ada untuk membuat beberapa data lagi dalam proses pelatihan model. *Image augmentation* memperluas variasi data dalam proses pelatihan yang lebih mendalam. Tahap terakhir pada persiapan data adalah membagi data menjadi dua yaitu data latih dan data validasi, dengan presentase 80% untuk data latih serta 20% untuk data validasi.

3.3 Pemodelan

Tahap pemodelan dilakukan dengan meingimplementasikan algoritma CNN atau *Convolutional Neural Network*. CNN dapat menganalisis visual gambar, mendeteksi suatu objek pada gambar yang merupakan vektor dengan dimensi tinggi yang melibatkan banyak parameter dalam melakukan pencirian jaringan. (Nugroho, Fenriana, & Arijanto, 2020). Ukuran gambar yang digunakan sebagai masukan adalah sebesar 150 x 150 *pixels*. Beberapa *layer* yang digunakan pada proses ini yaitu *convolutional layer* untuk mendeteksi keberadaan sekumpulan fitur pada gambar yang diterima sebagai input, *pooling layer* untuk mengurangi ukuran gambar dengan tetap mempertahankan karakteristik penting dari gambar tersebut, *flatten layer* untuk mengubah data menjadi *array* satu dimensi untuk memasukkannya ke *layer* berikutnya, dan *dense layer* yang digunakan untuk mengklasifikasikan citra berdasarkan keluaran dari *convolutional layer*.

3.4 Evaluasi

Proses pelatihan dilakukan pada beberapa model dengan konfigurasi yang berbeda. Model dikontrol untuk menghindari *overfitting* dengan menggunakan salah satu cara yaitu *hyperparameter tuning*. Cara ini digunakan untuk menemukan kumpulan *hyperparameter* terbaik yang mungkin untuk membangun model dari kumpulan data yang ada. Beberapa *hyperparameter* yang dapat diubah adalah jumlah neuron, fungsi aktivasi, *optimizer*, *learning rate*, ukuran *batch*, dan *epoch*. Setelah melakukan proses tersebut, dihasilkan satu model dengan konsistensi skor akurasi yang tinggi yaitu 89.5% pada data latih, dan 80.5% pada data validasi. Model ini kemudian disimpan dan digunakan pada proses selanjutnya yaitu *deployment*.

3.5 Finalisasi (Deployment)

Pada tahap finalisasi proyek *machine learning* dilakukan tahap *deployment*, yaitu proses mengambil model *machine learning* yang sudah dilatih dan membuat prediksinya tersedia untuk pengguna atau sistem lain. Pada proyek ini, model *machine learning* yang telah dibangun kemudian melalui tahap *deployment* ke dalam bentuk *website*. *Tools* yang digunakan pada proses *deployment* ini adalah Python Flask untuk integrasi model ke dalam bentuk *website* (HTML) dan Heroku sebagai *Platform as a Service* (PaaS).

Setelah tahap *deployment* dilakukan, proyek *machine learning* prediksi tuberkulosis ini sudah dapat diakses secara daring atau *online* melalui URL <https://classification-tuberculosis.herokuapp.com/>. Berikut adalah beberapa fitur dan tampilan pada *website* proyek *machine learning* yang sudah dibuat.

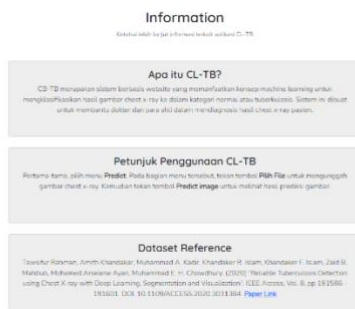
3.5.1 Fitur Home



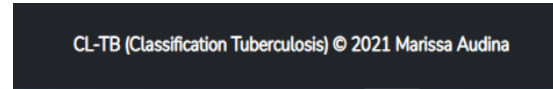
Gambar 3.1 Tampilan Fitur Home



Gambar 3.2 Tampilan Fitur Predict



Gambar 3.3 Tampilan Fitur *Info*



Gambar 3.4 Tampilan Fitur *About*

Pada fitur *Home*, terdapat tampilan awal saat *website* diakses. Pada bagian paling atas, terdapat navigasi yang berisi beberapa bagian yaitu *Home*, *Predict*, *Information*, dan *About*. Masing-masing bagian tersebut dapat ditekan dan mengarahkan *user* ke fitur-fitur yang dimaksud. Fitur ini menampilkan nama aplikasi yaitu “*Classification Tuberculosis (CL-TB)*”. Pada fitur ini dijelaskan juga informasi singkat mengenai aplikasi CL-TB. Tombol “*Start Predict*” dapat ditekan untuk langsung diarahkan ke fitur *Predict*.

3.5.2 Fitur *Predict*

Fitur *Predict* adalah fitur utama dari aplikasi proyek *machine learning* ini. Fitur *Predict* digunakan untuk memprediksikan data gambar *chest x-ray* dan menampilkan hasil prediksi dari gambar tersebut. *User* terlebih dahulu mengunggah file data gambar yang ingin diprediksi, kemudian gambar akan diprediksi menggunakan model yang sudah dilatih sebelumnya, dan menampilkan hasil prediksi gambar *chest x-ray* tersebut kepada *user*.

3.5.3 Fitur *Info*

Pada fitur *Info*, terdapat beberapa informasi terkait aplikasi CL-TB. Informasi yang pertama adalah tentang aplikasi CL-TB. Informasi selanjutnya adalah informasi mengenai petunjuk penggunaan aplikasi CL-TB. Kemudian bagian yang terakhir adalah informasi mengenai sumber referensi dataset yang digunakan untuk membangun model *machine learning* dalam proses klasifikasi data gambar *chest x-ray*.

3.5.4 Fitur *About*

Pada fitur ini ditampilkan *footer* untuk memberi informasi mengenai *copyright*, nama pemilik dan tahun pembuatan aplikasi.

3.6 Pengujian Sistem

Pengujian sistem dilakukan dengan *black box testing*. Pengujian ini berfokus pada fungsionalitas sistem tanpa melihat struktur internal atau kode program yang dibuat. Berikut merupakan hasil dari pengujian *black box testing*.

Tabel 3.1 Pengujian Sistem dengan *Black Box*

No	Skenario Pengujian	Luaran yang diharapkan	Status
1	<i>Home</i>	Menampilkan navigasi, nama dan deskripsi singkat aplikasi atau sistem.	Valid
2	<i>Predict</i>	Pengguna dapat mengunggah gambar <i>chest x-ray</i> dan sistem dapat	Valid

		menampilkan hasil prediksi yang sesuai (normal atau tuberkulosis).	
3	<i>Info</i>	Menampilkan informasi mengenai aplikasi atau sistem, petunjuk penggunaan, dan referensi dataset.	Valid
4	<i>About</i>	Menampilkan informasi mengenai <i>copyright</i> , nama pemilik dan tahun pembuatan aplikasi.	Valid

4. KESIMPULAN

Proyek akhir dari kegiatan MBKM Studi Independen Bersertifikat Dicoding 2021 yaitu pengembangan proyek *machine learning* telah dilakukan dan menghasilkan aplikasi berbasis *website* bernama CL-TB yang dapat memprediksikan data gambar *chest x-ray* yang mengidap tuberkulosis ataupun tidak. Fitur-fitur pada aplikasi ini juga telah berfungsi dengan baik dan benar sesuai dengan hasil pengujian. Adanya aplikasi ini dapat menjadi pendukung keputusan dokter serta tenaga ahli kesehatan dalam mendiagnosis pasien pengidap tuberkulosis.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada dosen pembimbing, serta seluruh pihak yang telah membantu kelancaran pembuatan proyek *machine learning* serta penyusunan jurnal pengabdian ini hingga selesai.

DAFTAR PUSTAKA

- Chomaerah, S. (2020). Program Pencegahan Dan Penanggulangan Tuberkulosis Di Puskesmas. *Higeia Journal of Public Health Research and Development*, 4(3), 398–410.
- Delrue, L., Gosselin, R., Ilsen, B., Van Landeghem, A., de Mey, J., & Duyck, P. (2011). *Difficulties in the Interpretation of Chest Radiography*. *June 2014*, 27–49. https://doi.org/10.1007/978-3-540-79942-9_2
- Dicoding. (2021). *Siapkan Kariermu jadi Developer Profesional bersama Dicoding dan Kampus Merdeka*. <https://www.dicoding.com/blog/kampusmerdeka-2022/>
- Febrianti, A. S., Sardjono, T. A., & Biomedik. (2020). *Klasifikasi Tumor Otak pada Citra Magnetic Resonance Image dengan Menggunakan Metode Support Vector Machine*. 9(1).
- Geneva:World Health Organization. (2021). *Global Tuberculosis Report*. Licence: CC BY-NC-SA 3.0 IGO.
- IBM Cloud Education. (2020). *Machine Learning*. <https://www.ibm.com/cloud/learn/machine-learning>
- Kementerian Kesehatan Republik Indonesia. (2021). *Dashboard TB*. <https://tbindonesia.or.id/pustaka-tbc/dashboard-tb/>
- Nayak, S. R., Nayak, D. R., Sinha, U., Arora, V., & Pachori, R. B. (2021). Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study. *Biomedical Signal Processing and Control*, 64(June 2020), 102365. <https://doi.org/10.1016/j.bspc.2020.102365>
- Nugroho, P. A., Fenriana, I., & Arijanto, R. (2020). Implementasi Deep Learning Menggunakan Convolutional Neural Network (Cnn) Pada Ekspresi Manusia. *Algor*, 2(1), 12–21.
- Rahman, T., Khandakar A., Kadir M. A., Islam K., R., Islam K. F., Mahbub Z. B., Ayari M. A., Chowdhury M. E. H. (2020) "Reliable Tuberculosis Detection using Chest X-ray with Deep Learning, Segmentation and Visualization". *IEEE Access*, Vol. 8, pp 191586 - 191601. DOI. 10.1109/ACCESS.2020.3031384.