

Pemetaan Proyeksi Penduduk Indonesia (Berdasarkan Jenis Kelamin) dengan *Machine Learning*

Nabhan Thoriq Ariyanto¹, Resa Setyawan², Rianto³, Vega Purwayoga⁴

^{1,2,3,4} Universitas Siliwangi, Tasikmalaya

Email : thoriqariyanto@gmail.com¹, resasetyawan29@gmail.com², rianto@unsil.ac.id³,
vega.purwayoga@unsil.ac.id⁴

Abstrak

Penggunaan Teknik *Machine Learning* telah menjadi sebuah metode yang vital dalam meramalkan jumlah penduduk di masa depan. Di Indonesia, proyeksi jumlah penduduk berdasarkan jenis kelamin memiliki dampak yang signifikan dalam merencanakan Pembangunan nasional. Penelitian ini mengusulkan suatu pendekatan yang memanfaatkan Teknik pemodelan *Machine Learning* untuk memproyeksikan jumlah penduduk Indonesia berdasarkan jenis kelamin. Dataset yang digunakan mencakup berbagai informasi seperti jumlah penduduk, provinsi, tahun, dan jenis kelamin. Proses pemodelan melibatkan tahapan-tahapan seperti persiapan data, pemodelan, dan evaluasi. Hasil proyeksi kemudian dinilai menggunakan berbagai metrik evaluasi kinerja model yang relevan.

Kata Kunci : *Jumlah penduduk, Jenis kelamin, Proyeksi, python, pembelajaran mesin*

Abstract

The use of Machine Learning techniques has become a vital method in predicting future population numbers. In Indonesia, population projections based on gender have a significant impact in planning national development. This research proposes an approach that utilizes Machine Learning modeling techniques to project the population of Indonesia based on gender. The dataset used includes various information such as population, province, year, and gender. The modeling process involves stages such as data preparation, model selection, model training, evaluation, and validation. The projection results are then assessed using various relevant model performance evaluation metrics.

Keyword : *Population, Gender, Projections, python, Machine Learning*

PENDAHULUAN

Dataset diperlukan untuk mengetahui sebuah informasi lebih spesifik secara angka sehingga program-program yang akan dilaksanakan dapat menjauhi kesalahan atau *error* yang lebih besar. Indonesia atau negara-negara lainnya pasti memiliki sebuah program pemerintah yang dilaksanakan baik itu untuk kebaikan negara atau masyarakat bangsa tersebut dan untuk melaksanakan program yang telah direncanakan diperlukan sebuah data yang mengetahui informasi dari

penduduk, dalam *paper* ini penulis mengambil dataset proyeksi jumlah penduduk yang didalamnya terdapat sebuah informasi mengenai provinsi dan jumlah penduduk tahun ketahun yang dibagi berdasarkan jenis kelamin yang mungkin diperlukan untuk program yang akan dilaksanakan oleh pemerintahan Indonesia. Perkiraan jumlah penduduk salah satu alat penting dalam merencanakan kemajuan suatu negara. Proyeksi jumlah penduduk merupakan salah satu alat penting dalam

merencanakan kemajuan suatu negara. Dengan memproyeksikan jumlah penduduk di masa depan, pemerintah bisa merancang kebijakan yang sesuai dalam sektor-sektor krusial seperti pendidikan, kesehatan, dan infrastruktur. Indonesia, yang memiliki populasi terbesar keempat di dunia, memperhitungkan proyeksi jumlah penduduk secara perencanaan pembangunan nasional. Proyeksi tersebut tidak hanya mempertimbangkan total populasi tetapi juga distribusi penduduk berdasarkan jenis kelamin. Pemodelan *Machine Learning* menawarkan pendekatan yang kuat dalam meramalkan jumlah penduduk di masa depan. (Sara Sheehan, 2016)

Indonesia, yang memiliki populasi terbesar keempat di dunia, memperhitungkan proyeksi jumlah penduduk secara perencanaan Pembangunan nasional. Proyeksi tersebut tidak hanya mempertimbangkan total populasi, tetapi juga distribusi penduduk berdasarkan jenis kelamin. Pemodelan *Machine Learning* menawarkan pendekatan yang kuat dalam meramalkan jumlah penduduk di masa mendatang (BPS, 2023).

PENELITIAN RELEVAN

Dalam penelitian ada beberapa aspek penting, yaitu:

Data demografis, Data ini akan dianalisis untuk memenuhi tren pertumbuhan penduduk Indonesia dari masa lalu hingga sekarang, serta untuk mengeluasi komposisi penduduk berdasarkan jenis kelamin. Faktor-faktor penyebab, Studi ini mempengaruhi pertumbuhan penduduk berdasarkan jenis kelamin, seperti Tingkat kelahiran, angka kematian, migrasi, dan factor-faktor sosial ekonomi lainnya.

Proyeksi masa depan, berdasarkan analisis data akan membuat proyeksi tentang bagaimana jumlah penduduk Indonesia akan berkembang di masa mendatang berdasarkan

jenis kelamin. Penerapan *Machine Learning* dalam proyeksi jumlah penduduk menawarkan berbagai manfaat, termasuk kemampuan untuk memproses data dalam volume besar dan mengidentifikasi pola yang tidak tampak secara langsung. Metode ini tidak hanya memperkuat kemampuan analitis, tetapi juga meningkatkan akurasi prediksi, yang pada akhirnya dapat mendukung kebijakan publik yang lebih responsif dan adaptif terhadap perubahan demografi.

Dengan demikian, studi kasus ini akan memberikan wawasan tentang bagaimana komposisi penduduk Indonesia berdasarkan jenis kelamin diperkirakan akan berubah seiring waktu, serta perubahan tersebut dalam perencanaan kebijakan dan Pembangunan sosial-ekonomi.

METODE PENELITIAN

Metode studi kasus Proyeksi Jumlah Penduduk Indonesia berdasarkan jenis kelamin menggunakan Pemodelan *Machine Learning* melibatkan sebagai berikut.

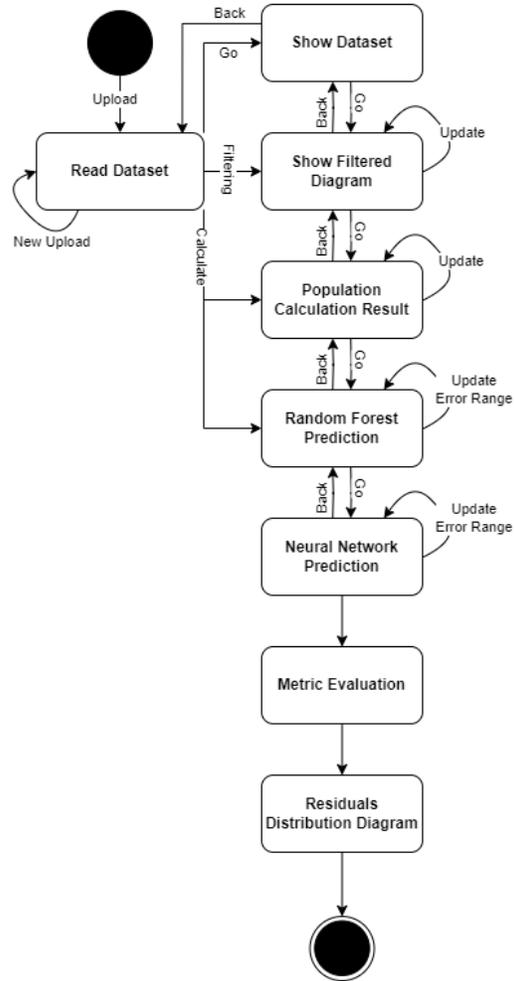
- Persiapan Data, Dalam melakukan proyeksi jumlah penduduk Indonesia berdasarkan jenis kelamin, kami menerapkan serangkaian langkah persiapan data yang sistematis untuk memastikan akurasi dan kualitas dataset yang digunakan. Data Penduduk Indonesia berdasarkan jenis kelamin dari tahun 2010 hingga 2035 yang diambil dari proyeksi jumlah penduduk Indonesia melalui Kaggle;
- Pemodelan *Machine Learning*, digunakan seperti *Random Forest* dan *Neural Network* untuk melakukan proyeksi jumlah penduduk berdasarkan jenis kelamin di masa mendatang. Pendekatan ini mencakup beberapa langkah utama dengan mengumpulkan dan membersihkan data dari Kaggle, mentransformasi data untuk memastikan

konsistensi skala, membagi dataset menjadi subset training dan testing serta mengevaluasi model; dan

- Evaluasi, kami menggunakan Evaluasi matrik untuk menilai kemampuan generalisasi model yang dilatih, serta untuk membandingkan berbagai model dalam memilih solusi optimal. Evaluasi matrik seperti akurasi sering digunakan, namun memiliki keterbatasan terutama dalam situasi ketidakseimbangan kelas. Oleh karena itu, pemilihan matrik evaluasi yang sesuai adalah langkah penting dalam mendapatkan model yang optimal. (Hossin, 2015)

HASIL DAN PEMBAHASAN

Untuk memaparkan dan menjelaskan dataset yang telah dipilih, kami menggunakan pemrograman Bahasa Python dengan bantuan GoogleCollabs yang tersedia di internet, karena Bahasa python merupakan Bahasa yang maju, mudah, dan mendukung banyak fitur seperti GUI atau AI sedangkan alasan menggunakan GoogleCollabs karena mudah diakses belum lagi merupakan salah satu compiler atau notebook bahasa python yang terkenal dan didalamnya sudah memiliki banyak library python sehingga mampu mendukung kita untuk melakukan penelitian ini dengan baik. Rencana yang akan dilaksanakan adalah mengikuti yang dipaparkan dalam Metode Penelitian secara tersusun dan rapi. Serta mengembangkan aplikasi dengan GUI sederhana yang dapat digunakan dalam Google Collab NoteBooks sehingga interactive dan dapat lebih mengefisienkan atau memudahkan dalam menggunakan aplikasi untuk menganalisa dataset tersebut.



Gambar 1. Diagram State untuk aplikasi yang dirancang

Kami membuat sebuah state diagram untuk memperjelas proses tahapan dan jalannya aplikasi yang telah direncanakan, terdapat beberapa proses sebelumnya untuk mencapai dari awal hingga final state.

Langkah pertama yang dilakukan adalah membaca dengan cara upload dataset dan menampilkan sebuah dataset tersebut yang dapat menghasilkan *Output*.

Isi Dataset:

| kode | provinsi | tahun | kelamin | jumlah |
|------|----------|-------|---------|--------|
| 11 | Aceh | 2010 | L | 2255.0 |
| 11 | Aceh | 2010 | P | 2268.1 |
| 11 | Aceh | 2015 | L | 2497.5 |
| 11 | Aceh | 2015 | P | 2504.5 |
| 11 | Aceh | 2020 | L | 2728.2 |
| 11 | Aceh | 2020 | P | 2731.7 |
| 11 | Aceh | 2025 | L | 2933.3 |
| 11 | Aceh | 2025 | P | 2936.7 |

Gambar 2. Menampilkan Isi dataset dalam bentuk tabel

Dengan menggunakan bantuan Library Pandas untuk membaca File csv dan IPython untuk menampilkan dataset dalam bentuk tabel yang rapih kita mampu memaparkan dan menganalisanya jauh lebih jelas dan mudah.

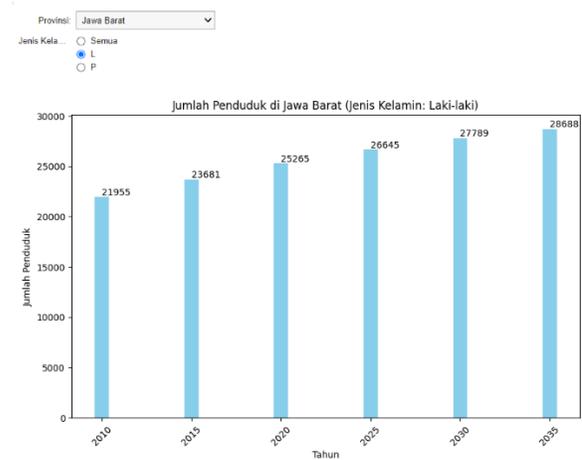
| | | | | |
|----|-------|------|---|--------|
| 94 | Papua | 2015 | P | 1488.2 |
| 94 | Papua | 2020 | L | 1802.2 |
| 94 | Papua | 2020 | P | 1633.2 |
| 94 | Papua | 2025 | L | 1931.4 |
| 94 | Papua | 2025 | P | 1770.3 |
| 94 | Papua | 2030 | L | 2044.3 |
| 94 | Papua | 2030 | P | 1895.1 |
| 94 | Papua | 2035 | L | 2139.1 |
| 94 | Papua | 2035 | P | 2005.5 |

Jumlah total baris dalam dataset: 396

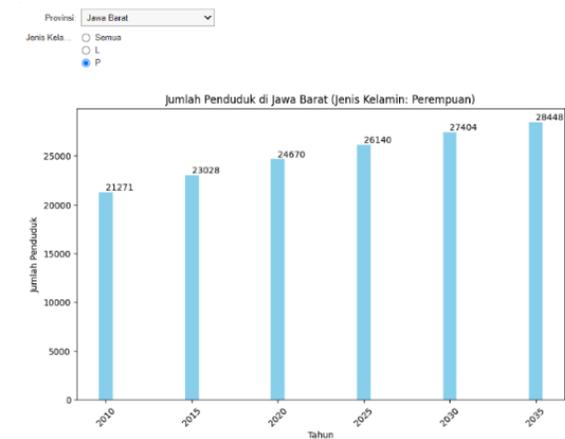
Gambar 3. Isi dataset serta total jumlah data (baris)

Isi dataset tersebut adalah sebanyak 396 baris yang isinya adalah sebuah proyeksi perkembangan populasi penduduk di setiap provinsi Indonesia. Untuk menampilkan proyeksi perkembangan penduduk dari tahun ke tahun yang lebih jelas demi keefisienan dalam menganalisa penulis membuat sebuah kode untuk menampilkan diagram untuk setiap provinsi dengan mengkombinasikan

library Pyplot dan Ipywidgets hasil yang diinginkanpun mampu tercapai.

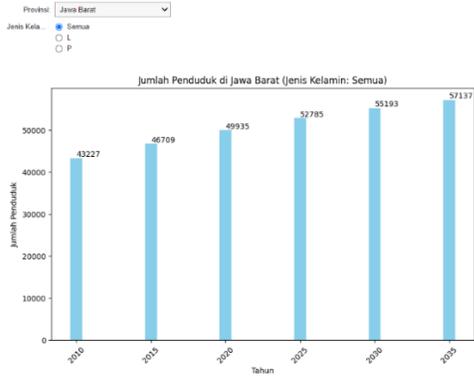


Gambar 4. Proyeksi Laki-laki provinsi Jawa Barat

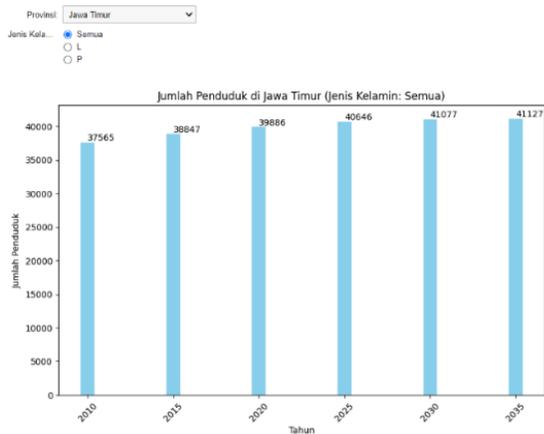


Gambar 5. Proyeksi perempuan provinsi Jawa Barat

Dapat dilihat dalam gambar diatas adalah proyeksi jumlah penduduk dari salah satu provinsi yang penulis ambil dari contoh yaitu provinsi jawa barat dapat dilihat perkembangan dari tahun ketahun berdasarkan gendernya. Jika diperhatikan secara seksama jumlah laki-laki lebih banyak daripada Perempuan di provinsi Jawa Barat.



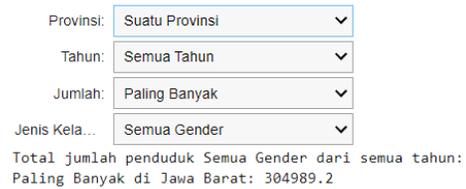
Gambar 6. Proyeksi jumlah penduduk Jawa Barat



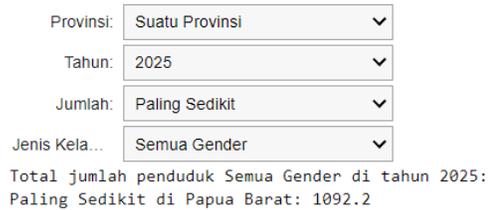
Gambar 7. Proyeksi jumlah penduduk Jawa Timur

Dalam diagram di atas, terlihat bahwa provinsi Jawa Barat memiliki proyeksi populasi total yang lebih banyak dibandingkan dengan Jawa Timur dari tahun ke tahun. Data ini menunjukkan pertumbuhan populasi yang lebih cepat di Jawa Barat dibandingkan dengan Jawa Timur. Untuk memperoleh data yang lebih rinci dan akurat, kami mengambil langkah lebih lanjut dengan membuat sebuah program untuk proses selanjutnya yang bertujuan menghitung dan menganalisis perbedaan populasi antara kedua provinsi tersebut. Dengan menggunakan program ini, data yang dihasilkan menjadi lebih terperinci,

memungkinkan perbandingan yang lebih mendalam antara pertumbuhan populasi di Jawa Barat dan Jawa Timur. Dalam proses ini memberikan sebuah nilai dari perhitungan dengan menggunakan proses *filtering* sebuah data yang membantu dalam memahami dinamika populasi di kedua provinsi tersebut. Hasil perhitungan dari program ini akan ditampilkan pada gambar berikutnya di bawah ini.



Gambar 8. Proses *filtering* dengan ketentuan ke-1



Gambar 9. Proses *filtering* dengan ketentuan ke-2

Disini Pengguna akan berinteraksi dengan sebuah antarmuka yang telah dirancang khusus untuk memungkinkan pemilihan berdasarkan berbagai kategori yang tersedia. Antarmuka ini mempermudah pengguna dalam memilah pilihan yang disediakan, memastikan bahwa setiap kategori dapat diakses dengan mudah. Dengan menggunakan fitur tersebut, pengguna dapat melakukan seleksi secara efisien dan menemukan apa yang mereka butuhkan sesuai dengan preferensi dan kebutuhannya.

Provinsi: Riau
Tahun: Semua Tahun
Jumlah: Paling Banyak
Jenis Kela... Perempuan

Total jumlah penduduk paling banyak di Riau adalah 4592.900 untuk Perempuan pada tahun 2035

Gambar 11. Proses *filtering* dengan ketentuan ke-3

Provinsi: Semua Provinsi
Tahun: Semua Tahun
Jumlah: Paling Banyak
Jenis Kela... Semua Gender

Total jumlah penduduk di Indonesia adalah 1651933.400 untuk semua gender pada semua tahun

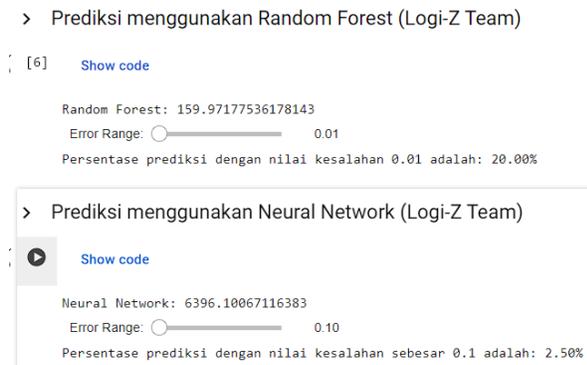
Gambar 10. Proses *filtering* dengan ketentuan ke-4

Penulis membuat program tersebut untuk semacam menyaring data berdasarkan tag ketentuan tertentu dan dapat menghasilkan *output* yang perlu diketahui atau diinginkan oleh *user*. Oleh karena itu, dalam program tersebut telah tersaring dan diketahui bahwa provinsi Jawa Barat memiliki jumlah populasi paling banyak dari seluruh provinsi Indonesia untuk Semua tahun (artinya total jumlah penduduk dari tahun 2020 sampai 2035) dan Paling sedikit adalah di Provinsi Papua Barat pada tahun 2025 dengan jumlah 1092 jiwa, sedangkan untuk seluruh provinsi dan semua tahun di Indonesia memiliki Total 1.651.933 jiwa untuk semua gender, Selanjutnya terdapat sebuah pilihan untuk memilih provinsi yang terdapat dalam dataset lalu program akan mulai membandingkan dan menghitung jumlah paling banyak atau paling sedikit dari suatu gender tertentu dan menampilkan total jumlah untuk tahun berapa provinsi tersebut memiliki populasi paling banyak begitu juga sebaliknya, jika dalam gambar dari contoh yang diambil Provinsi Riau dengan gender Perempuan memiliki jumlah penduduk 4592 jiwa pada tahun 2035 yang merupakan jumlah proyeksi jumlah penduduk terbanyak dari semua tahun. Dari analisis hasil program tersebut

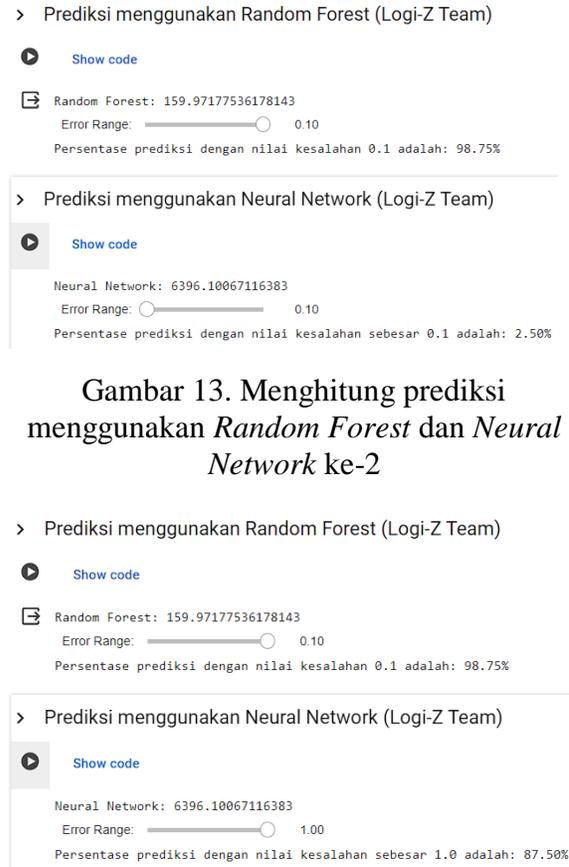
dan mengambil data dari proyeksi dataset yang digunakan seluruh provinsi di Indonesia populasi penduduk Indonesia memiliki perkembangan yang bertambah dalam setiap 5 tahun untuk setiap provinsi pemegang populasi terbanyak dimiliki oleh provinsi Jawa Barat dan paling sedikit oleh Papua Barat.

Langkah berikutnya adalah melakukan pemodelan dari *Machine Learning* dengan menggunakan *Random Forest* untuk mengkalkulasi jumlah populasi penduduk dimasa yang akan datang. Tentunya disini kami masih menggunakan Bahasa pemrograman python dan GoogleCollabs sebagai compilernya. Salah satu alasan kami menggunakan *Random Forest* adalah karena metode tersebut merupakan pendekatan pemodelan ensemble nonparametrik yang membentuk "hutan" dari pohon klasifikasi atau regresi individu. Metode ini memperbaiki teknik bagging dengan menggunakan yang terbaik dari pilihan acak prediktor pada setiap node di setiap pohon (Forrest R. Stevens, 2015). Jadi, *Random Forest* adalah sebuah pendekatan pemodelan ensemble dan nonparametrik yang membangun sebuah "hutan" dari kumpulan pohon keputusan klasifikasi atau regresi.

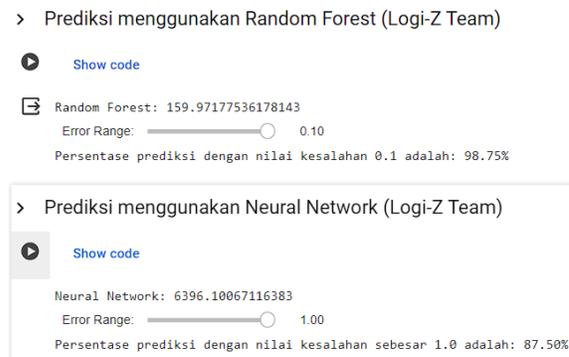
Metode ini meningkatkan performa dari bagging dengan menggunakan prediktor terbaik dari suatu seleksi acak pada setiap simpul (node) di setiap pohon. Dalam konteks perkembangan populasi penduduk, *Random Forest* dapat diumpamakan seperti proses pengambilan keputusan yang melibatkan banyak ahli yang berbeda. Setiap pohon keputusan dalam hutan tersebut mewakili salah satu ahli, dan pada setiap langkah dalam pengambilan keputusan, *Random Forest* akan mempertimbangkan pandangan dari beberapa ahli tersebut secara acak. Dengan demikian, *Random Forest* dapat menghasilkan prediksi yang lebih stabil dan akurat terhadap perkembangan populasi penduduk dengan memanfaatkan berbagai faktor yang relevan. Dibawah ini adalah aplikasi yang telah dibuat untuk menghitung prediksi menggunakan *Random Forest* dan *Neural Network* sesuai dengan metode penelitian yang akan kita gunakan.



Gambar 12. Menghitung prediksi menggunakan *Random Forest* dan *Neural Network* ke-1



Gambar 13. Menghitung prediksi menggunakan *Random Forest* dan *Neural Network* ke-2



Gambar 14. Menghitung prediksi menggunakan *Random Forest* dan *Neural Network* ke-3

Pertama dalam pembuatan aplikasi untuk memprediksi nilai, *developer* hanya membuat perhitungan sederhana yang menampilkan nilai prediksi RMSE (*Root Mean Square Error*) karena rasanya angka tersebut sedikit sulit untuk dimengerti oleh orang awam maka pengembang membuat sebuah perhitungan agar nilai tersebut terpaparkan dalam persentase setelah melaksanakan beberapa test uji dan penelitian yang akhirnya dapat menemukan titik cerah dan dapat menampilkan prediksi tersebut dalam sebuah persentase dan dikembangkan lebih jauh sampai bisa memodifikasi nilai kesalahan menggunakan UI. Pengembang disini membuat semacam *slider* untuk mengatur atau mengubah *error value* dengan lebih mudah tanpa harus

mengubah dari kodingan atau mengetiknya secara satu-persatu sesuai dengan salah satu visi kita juga untuk membuat GUI sederhana yang interactive dengan pengguna. Penulis menampilkan 3 contoh kasus untuk membandingkan metode *Neural Network* dan *Random Forest*, dapat dilihat hasil prediksi kedua metode tersebut memiliki nilai hasil rentang yang cukup jauh dan setelah memodifikasi kodingan berkali-kali seperti halnya hasil perhitungan tersebut cocok untuk model yang sedang kami gunakan dikarenakan tidak ada *error* sama sekali dan perhitungan sudah sesuai dengan yang digunakan secara umum dalam penghitungan RF dan NN. Seperti yang diketahui semakin kecil *error value* yang digunakan semakin akurat kemungkinan prediksi yang didapatkan begitu juga sebaliknya, jika *error value* yang dimiliki itu besar maka akan semakin tidak akurat nilai prediksi yang dihitung, dalam perhitungan prediksi tersebut pengembang menggunakan *range error value* yang dasarnya berbeda contohnya *Random Forest* memiliki nilai kesalahan minimum di 0.01 sedangkan NN di 0.1 untuk nilai maksimum RF di 0.1 dan NN di 1.0 mengapa demikian? Selama testing aplikasi dan mengetest perhitungan prediksi model RF menghasilkan nilai prediksi yang lebih besar dibandingkan NN contohnya jika *error value* diatas 0.1 untuk model RF dia akan menghasilkan prediksi presentase keberhasilan 100% yang Dimana hasil tersebut bisa dikatakan tidak berguna karena semakin besar nilai *error value* itu jadi tidak akurat jadi pengembang membuat *error value* maksimum di 0.1 untuk RF dan bisa diperkecil hingga 0.01 yang hasil prediksi keberhasilannya adalah 20% sedangkan untuk NN dengan nilai kesalahan maksimum sebesar 1.0 dan dapat diperkecil hingga 0.1 yang menampilkan nilai hasil prediksi keberhasilan sebesar 2.5% pengembang sengaja membuat *error value* dapat dimodifikasi lebih dari 0.1 karena untuk

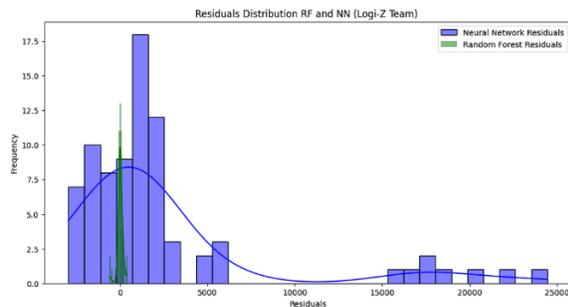
model NN jika *error value* lebih kecil dari 0.1 akan memungkinkan munculnya nilai prediksi 0% yang dimana hasil tersebut tidak dapat diperhitungkan juga atau sia-sia karena tidak dapat dianalisa. Dari hasil test uji coba tersebut kami yakin bahwa model RF lebih cocok digunakan dari NN dalam kasus ini karena bisa dikatakan lebih akurat dan baik oleh sebab RF masih mampu menampilkan persentase yang lebih tinggi dengan *error value* yang lebih kecil dari 0.1 sedangkan untuk NN perhitungan persentase nilai kesalahan harus diatas 0.1 semakin kecil nilai kesalahan maka semakin akurat prediksinya oleh karena itu model RF lebih cocok untuk dataset yang sedang digunakan. Selanjutnya, sebagai penutup dan langkah terakhir kami mengevaluasi menggunakan metrik evaluasi untuk membandingkan kedua model tersebut mana yang lebih baik atau seberapa efisien model yang digunakan.

Tabel 1. Menghitung akurasi prediksi menggunakan evaluasi matrik

| Model | R2 Score |
|----------------|-----------|
| Neural Network | -0.151184 |
| Random Forest | 0.999280 |

Dapat dilihat dalam tabel 1. Merupakan hasil penjumlahan dari program yang dibentuk dalam tabel dan menampilkan nilai koefisien determinasi (*R-squared*) untuk kedua model tersebut, dapat dilihat bahwa hasil evaluasi metrik untuk model NN adalah negatif sedangkan RF positif apa artinya? Model NN yang kami gunakan tidak sesuai untuk dataset yang kami gunakan. Biasanya nilai negatif untuk koefisien determinasi sering terjadi dalam kasus di mana model benar-benar tidak cocok dengan data dan prediksinya tidak lebih baik dari *random*. Dalam praktiknya, nilai negatif untuk metrik

evaluasi menunjukkan bahwa model tidak memiliki nilai prediktif yang berguna dan perlu direvisi atau dibuat kembali. Sedangkan sebaliknya untuk model RF yang kami gunakan, ia menampilkan nilai positif sebesar 0.999 sangat mendekati satu dan itu menunjukkan bahwa model RF sangat baik dan efisien untuk pengolahan data dan prediksi nilai dari dataset yang kami gunakan. Oleh karena itu, kami memutuskan bahwa model RF dalam konteks dataset yang kami gunakan lebih baik daripada model NN dengan R2-score bernilai negatif. Kami juga mencoba membandingkan dengan metode distribution residuals untuk memvisualisasi dalam diagram.



Gambar 15. Hasil perbandingan menggunakan distribusi residuals

Distribusi residuals menggambarkan perbedaan antara nilai aktual dari variabel yang diprediksi dan nilai yang diprediksi oleh model. Residuals adalah selisih antara nilai sebenarnya dari data dan nilai yang diprediksi oleh model. Distribusi residuals menunjukkan seberapa sering nilai residuals tersebut muncul pada berbagai rentang. Pertama, kita menghitung residual untuk masing-masing model (NN dan RF) dengan cara mengurangkan nilai aktual (y_{test}) dengan nilai prediksi yang dihasilkan oleh model. Selanjutnya, kita menggunakan histogram untuk menampilkan distribusi dari residuals. Histogram digunakan untuk menunjukkan sebaran frekuensi dari nilai-nilai residuals. Dalam histogram ini, sumbu x mewakili nilai residuals sedangkan sumbu y

mewakili frekuensi kemunculan nilai residuals tersebut dan yang terakhir Histogram tersebut memperlihatkan bagaimana residuals dari masing-masing model tersebar. Jika distribusi residuals berpusat di sekitar 0 dan terdistribusi secara simetris, itu menandakan bahwa model memiliki kinerja yang baik dalam memprediksi data. Namun, jika distribusi residuals cenderung miring atau tidak berpusat di sekitar 0, itu bisa menandakan adanya bias atau ketidaksempurnaan dalam model. Intinya semakin dekat dengan 0 maka semakin baik kinerja dari model yang dibuat jika dalam gambar dapat dilihat perbedaan RF dan NN itu sangat beda jauh yang membuktikan bahwa model NN itu sangat tidak efisien daripada RF dalam kasus penelitian yang kita angkat.

SIMPULAN

Penggunaan teknolgi pemodelan *Machine Learning* dapat menjadi sebuah sarana yang efisien dalam jumlah penduduk Indonesia dengan mempertimbangkan perbedaan jenis kelamin. Model yang akurat mampu mendukung pemerintah dalam Menyusun kebijakan yang lebih efektif dalam mengantisipasi perubahan demografis yang akan terjadi di masa depan. Dalam penelitian ini juga untuk perhitungan prediksi proyeksi pertumbuhan penduduk terbukti bahwa model *Random Forest* lebih efisien, berkinerja lebih baik dan efektif dari pada model *Neural Network*.

DAFTAR PUSTAKA

Annahl, M. A. (23. 10 2020). Performance Management in Indonesia Zakat Institutions: Balanced Scorecard and Fuzzy DEMATEL Framework. (2020: *The 4th International Conference of Zakat (ICONZ) Proceeding*). Surabaya. doi:<https://doi.org/10.37706/iconz.2020.2>

- Arnolus Juantri E. Oktavianus, L. N. (2023). Pemanfaatan Artificial Intelligence pada Pembelajaran dan Asesmen di Era Digitalisasi. *JURNAL KRIDATAMA SAINS DAN TEKNOLOGI*, 5(2), 473-484. doi:<https://doi.org/10.48550/arXiv.1711.09846>
- Elsa G. Guillot, M. K. (2013). Climate Change Influenced Female Population Sizes Through Time Across the Indonesian Archipelago. *Human Biology*, 135-152. doi:<https://doi.org/10.3378/027.085.0306>
- Emma V. A. Sylvester, P. B. (2017, July 11). Applications of *Random Forest* feature selection for fine- scale genetic population assignment. *Evolutionary Applications*, 11, 153-165. doi:<https://doi.org/10.1111/eva.12524>
- Forrest R. Stevens, A. E. (2015). *Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data*. Northwestern University. Evanston: Plos One. doi:<https://doi.org/10.1371/journal.pone.0107042>
- Hossin, M. S. (2015). A Review on Evaluation Metrics For Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2). Von <https://www.airconline.com/ijdkp/V5N2/5215ijdkp01.pdf>
- Juanmei Liu, Z.-H. T. (2013). Artificial Neural Network models for prediction of cardiovascular autonomic dysfunction in general Chinese population. *BMC Medical Informatics and Decision Making*, 13. Von <http://www.biomedcentral.com/1472-6947/13/80>
- Max Jaderberg, V. D. (2017). *Population Based Training of Neural Networks*. arXiv.
- McDonald, P. (2014). A Population Projection for Indonesia, 2010–2035. *Bulletin of Indonesian Economic Studies*, 123-129. doi:<https://doi.org/10.1080/00074918.2014.896240>
- Ni Komang Arista Dewi, a. L. (2020). PEMETAAN BENTUK DAN PENCEGAHAN PENIPUAN E-COMMERCE. *E-Jurnal Ekonomi dan Bisnis Universitas Udayana* 9.9, 851-878.
- Panggabean, M. (2020). Faktor-Faktor Yang Mempengaruhi Dependency Ratio di Indonesia. *Prosiding Seminar Akademik Tahunan Ilmu Ekonomi dan Studi Pembangunan 2020*, 371-387.
- Rémi Abgrall, É. L. (2022). Relaxation Deferred Correction Methods and their Applications to Residual Distribution Schemes. *SMAI Journal of Computational Mathematics*, 8, 125-160. doi:<https://doi.org/10.5802/smai-jcm.82>
- Roma Akbar Iswara, E. S. (2018). Sistem Pendukung Keputusan Untuk penentuan mustahik (Penerima Zakat) Menggunakan Metode Fuzzy AHP (F-AHP). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1306-1312.
- Sara Sheehan, Y. S. (28. March 2016). *Deep Learning for Population Genetic Inference*. Rutgers University. PLOS COMPUTATIONAL BIOLOGY. doi:<https://doi.org/10.1371/journal.pcbi.1004845>
- TAFHAMIN, D. A. (2020). *APLIKASI SOFTWARE SPECTRUM UNTUK GAMBARAN PROYEKSI PENDUDUK KOTA SURABAYA TAHUN 2018-2028*. AIRLANGGA University, Surabaya. Von <https://media.neliti.com/media/publicatio>

ns/50568-ID-proyeksi-penduduk-indonesia-per-provinsi-menurut-kelompok-umur-dan-jenis-kelamin.pdf

Yang Zhao, F. C. (2012). Correction for population stratification in *Random Forest* analysis. *International Journal of Epidemiology*, 41(6), 1798–1806. doi:<https://doi.org/10.1093/ije/dys183>

Yudistira, N. (2018). Peran Big Data dan Deep Learning untuk Menyelesaikan Permasalahan Secara Komprehensif. *EXPERT Jurnal Manajemen Sistem Informasi dan Teknologi*, 11, 79-89. doi:<http://dx.doi.org/10.36448/expert.v11i2.2063>