



**IMPLEMENTASI METODE *RANDOM FOREST* UNTUK  
MEMPREDIKSI INDEKS PEMBANGUNAN MANUSIA DI PROVINSI  
JAWA TENGAH**

**Tugas Akhir**

diajukan sebagai salah satu syarat untuk memperoleh gelar  
Ahli Madya Program Studi Statistika Terapan dan Komputasi

oleh

Dian Ayu Puspitasari

4112320036

**PROGRAM STUDI STATISTIKA TERAPAN DAN KOMPUTASI  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI SEMARANG, 2023**



**IMPLEMENTASI METODE *RANDOM FOREST* UNTUK  
MEMPREDIKSI INDEKS PEMBANGUNAN MANUSIA DI PROVINSI  
JAWA TENGAH**

**Tugas Akhir**

diajukan sebagai salah satu syarat untuk memperoleh gelar  
Ahli Madya Program Studi Statistika Terapan dan Komputasi

oleh

Dian Ayu Puspitasari

4112320036

**PROGRAM STUDI STATISTIKA TERAPAN DAN KOMPUTASI  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI SEMARANG, 2023**

## PERSETUJUAN PEMBIMBING

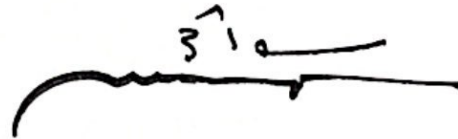
Tugas Akhir yang berjudul “Implementasi Metode *Random Forest* untuk memprediksi Indeks Pembangunan Manusia di Provinsi Jawa Tengah” yang disusun oleh

Nama : Dian Ayu Puspitasari  
NIM : 4112320036  
Prodi : Statistika Terapan dan Komputasi

telah disetujui untuk diajukan ke sidang ujian tugas akhir.

Semarang, 16 September 2023

Pembimbing



Dr. Walid, S. Pd., M.Si.

NIP 197408192001121001

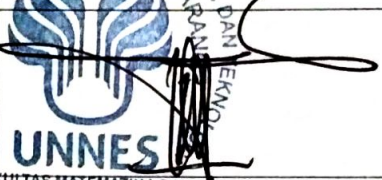

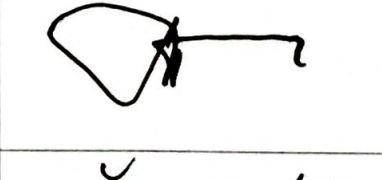
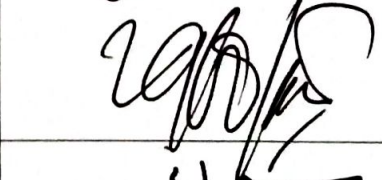
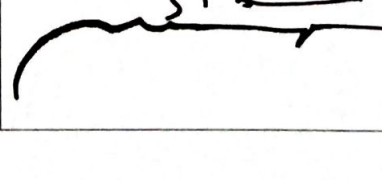
## PENGESAHAN PENGUJI

Tugas Akhir berjudul “Implementasi Metode *Random Forest* untuk Memprediksi Indeks Pembangunan Manusia di Provinsi Jawa Tengah” yang disusun oleh

Nama : Dian Ayu Puspitasari  
NIM : 4112320036  
Prodi : Statistika Terapan dan Komputasi

Telah dipertahankan dalam ujian tugas akhir pada hari Rabu, 11 Oktober 2023

Tim Penguji

<b>Ketua Panitia</b> Prof. Dr. Edy Cahyono, M.Si. NIP 196412051990021001	 UNNES FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
<b>Sekretaris Panitia</b> Dr. Putriaji Hendikawati, S.Si., M.Pd., M.Sc. NIP 198208182006042001	
<b>Penguji 1</b> Dr. Drs. Sugiman, M.Si. NIP 196401111989011001	
<b>Penguji 2</b> Dr. Iqbal Kharisudin, M.Sc. NIP 197908052005011003	
<b>Penguji 3/Pembimbing</b> Dr. Walid, S. Pd., M.Si. NIP 197408192001121001	

## PERNYATAAN

Tugas Akhir yang ditulis berjudul “Implementasi Metode *Random Forest* untuk Memprediksi Indeks Pembangunan Manusia di Provinsi Jawa Tengah” merupakan karya ilmiah asli. Bukan hasil plagiasi dari karya ilmiah orang lain atau pengutipan dengan cara-cara yang tidak sesuai dalam etika keilmuan yang berlaku baik sebagian atau seluruhnya. Pendapat atau temuan orang lain yang terdapat dalam Tugas Akhir ini telah dikutip atau dirujuk berdasarkan kode etik ilmiah.

Semarang, 11 Oktober 2023

Yang Menyatakan



Dian Ayu Puspitasari

NIM 4112320036

## MOTTO DAN PERSEMBAHAN

### MOTTO

Bersyukur dengan apa yang terjadi dalam hidup adalah hal baik.

Kita dilahirkan untuk menjadi nyata bukan menjadi sempurna. (Suga)

### PERSEMBAHAN

*Alhamdulillah rabbil 'alamin* atas izin dan Ridha ALLAH SWT, saya ingin mempersembahkan karya tulis ini untuk kedua orang tua saya Bapak Supadmo dan Ibu Sulastri yang selalu tiada hentinya mendoakan serta memberi dukungan supaya anaknya diberikan kemudahan maupun kelancaran dalam mengenyam pendidikan di bangku perkuliahan ini. Semangat dan senyum mereka menjadi motivasi saya untuk terus belajar dan berkembang sehingga sampai pada tahap ini. Selain itu, kepada kakak saya dan keluarga kecilnya yang selalu membantu, mendoakan, dan memberi dukungan selama proses perkuliahan saya.

Tidak lupa juga saya ucapkan terimakasih kepada para sahabat Nadia, Zannuba, Iin, Zuhrotu, Feny, dan Adella yang telah kebersamai selama tiga tahun ini. Suka maupun duka yang telah kita lewati bersama selama ini, saya berharap semoga selalu sehat dan semoga segala hal hal baik selalu menyertai kalian. Untuk teman-teman seperjuangan D3 Staterkom'20 terimakasih sudah berjuang bersama hingga kita dapat berada di titik ini dan selalu mengingatkan untuk dapat menyelesaikan kuliah dengan baik. Tidak lupa juga saya persembahkan karya tulis ini untuk diri saya sendiri yang sudah berjuang dan mampu bertahan sampai sejauh ini.

## ABSTRAK

Puspitasari, Dian Ayu (2023), Implementasi Metode *Random Forest* untuk Memprediksi Indeks Pembangunan Manusia di Provinsi Jawa Tengah. Tugas Akhir. Statistika Terapan dan Komputasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Dr. Walid, S. Pd., M.Si.

**Kata kunci:** IPM, *Random Forest*, dan *Decision Tree*

IPM digunakan sebagai alat untuk mengukur keberhasilan pembangunan manusia di suatu wilayah. Status IPM di Provinsi Jawa Tengah termasuk ke dalam kategori tinggi. Serta tingkat penyebaran kabupaten/kota status nilai IPM tergolong bervariasi. Oleh karena itu, upaya Pemerintah Jawa Tengah dalam menanggulangi ketidaksamaan tingkat IPM di kabupaten/kota maka dilakukan prediksi dengan metode regresi *random forest* dan *decision tree* yang tepat. Penelitian ini bertujuan untuk mengetahui hasil perbandingan dan implementasi metode *random forest regression* dan *decision tree regression* serta faktor yang mempengaruhi IPM di Jawa Tengah dengan bantuan *software python*.

Kedua metode penelitian yang digunakan pada penelitian ini merupakan algoritma *machine learning* yang memiliki dua fungsi penyelesaian masalah yaitu regresi dan klasifikasi. Metode regresi digunakan ketika variabel dependen berskala kontinu atau numerik. Pengukuran performa dan kualitas model prediksi pada penelitian ini menggunakan metrik evaluasi R-Square, MAPE, RMSE. Selain itu, untuk menentukan seberapa penting atau berpengaruhnya suatu variabel terhadap hasil prediksi model dapat dilihat dari nilai *variable importance*. Penelitian ini menggunakan data IPM di Provinsi Jawa Tengah tahun 2022 berdasarkan kabupaten/kota. Variabel yang digunakan yaitu usia harapan hidup, harapan lama sekolah, rata-rata lama sekolah, dan pengeluaran perkapita yang disesuaikan.

Hasil evaluasi model yang diperoleh dari nilai metrik evaluasi didapatkan nilai R-Square sebesar sebesar 0,953, MAPE sebesar 0,657 atau 65,7% serta RMSE sebesar 0,881 untuk metode *random forest*. Serta untuk metode *decision tree* diperoleh nilai R-Square sebesar 0,843, MAPE sebesar 1,544 atau 154,4% serta RMSE sebesar 1,610. Nilai R-Square yang tinggi menjelaskan variasi data serta untuk MAPE dan RMSE memiliki nilai yang relatif rendah menunjukkan model berkinerja baik dalam memprediksi nilai aktual. Oleh karena itu, metode terbaik yang dapat digunakan untuk memprediksi data IPM di Provinsi Jawa Tengah yaitu menggunakan metode *random forest regression*. Selain itu, Variabel dengan kepentingan terbesar yang mampu mempengaruhi nilai prediksi IPM adalah variabel rata-rata lama sekolah dengan nilai sebesar 0,972. Artinya bahwa variabel rata-rata lama sekolah memiliki dampak yang signifikan dalam menentukan nilai IPM, semakin tinggi nilai variabel rata-rata lama sekolah maka semakin tinggi pula kemungkinan nilai prediksi IPM.

## PRAKATA

Segala puji dan Syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik dan lancar. Tugas akhir yang berjudul “Implementasi Metode *Random Forest* dan *Decision Tree* untuk Memprediksi Indeks Pembangunan Manusia di Provinsi Jawa Tengah” disusun sebagai salah satu syarat untuk memperoleh gelar Diploma (D3) pada Program Studi Statistika Terapan dan Komputasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.

Diselesaikannya Tugas Akhir ini tidak lepas dari bantuan, bimbingan, dan doa dari berbagai pihak. Oleh karena itu, di kesempatan kali ini penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada yang terhormat:

1. Prof. Dr. S Martono, M.Si., selaku Rektor Universitas Negeri Semarang.
2. Prof. Dr. Edy Cahyono, M.Si., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.
3. Dr. Putriaji Hendikawati S.Si., M.Pd., M.Sc., selaku Koordinator Program Studi Statistika Terapan dan Komputasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.
4. Dr. Walid, S.Pd., M.Si., sebagai Dosen Pembimbing yang telah memberikan bimbingan, pengarahan, dan motivasi kepada penulis dalam penyusunan Tugas Akhir ini.
5. Bapak dan Ibu Dosen Program Studi Statistika Terapan dan Komputasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.
6. Kedua orang tua tercinta dan kakak yang senantiasa selalu mendoakan, memberi semangat, dan memberi dukungan maupun motivasi dalam penyusunan Tugas Akhir ini.
7. Berbagai pihak yang telah membantu dan tidak dapat penulis sebutkan satu persatu.
8. Diri saya sendiri, yang sudah berproses dan berjuang selama tiga tahun ini terima kasih karena sudah memberikan yang terbaik.

Semarang, 5 September 2023



## DAFTAR ISI

COVER .....	i
PERSETUJUAN PEMBIMBING .....	ii
PENGESAHAN PENGUJI .....	iii
PERNYATAAN.....	iv
MOTTO DAN PERSEMBAHAN.....	v
ABSTRAK .....	vi
PRAKATA .....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL .....	xi
<b>BAB I PENDAHULUAN</b> .....	1
<b>1.1 Latar Belakang</b> .....	1
<b>1.2 Rumusan Masalah</b> .....	4
<b>1.3 Tujuan Penelitian</b> .....	5
<b>1.4 Batasan Masalah</b> .....	5
<b>1.5 Manfaat Penelitian</b> .....	5
<b>1.6 Keaslian Penelitian</b> .....	6
<b>BAB II KAJIAN PUSTAKA</b> .....	10
<b>2.1 Tinjauan Pustaka</b> .....	10
<b>2.2 Landasan Teori</b> .....	13
<b>2.2.1 Data Mining</b> .....	13
<b>2.2.2 Machine learning</b> .....	15
<b>2.2.3 CART (Classification and Regression Tree)</b> .....	17
<b>2.2.4 Bootstrap Aggregating (Bagging)</b> .....	18
<b>2.2.5 Decision Tree (Pohon Keputusan)</b> .....	19
<b>2.2.6 Random forest (Hutan Acak)</b> .....	20
<b>2.2.7 Variable Importance</b> .....	23
<b>2.2.8 Metrik Evaluasi</b> .....	24
<b>2.2.9 Python</b> .....	26
<b>2.2.10 Google collaboratory</b> .....	27
<b>2.2.11 Indeks Pembangunan Manusia</b> .....	27
<b>2.3 Kerangka Berpikir</b> .....	30

<b>BAB III METODE PENELITIAN .....</b>	<b>32</b>
<b>3.1 Pendekatan dan Desain Penelitian.....</b>	<b>32</b>
<b>3.2 Fokus Penelitian .....</b>	<b>32</b>
<b>3.3 Variabel Penelitian .....</b>	<b>32</b>
<b>3.4 Data dan Sumber Data.....</b>	<b>34</b>
<b>3.5 Teknik Pengumpulan Data.....</b>	<b>34</b>
<b>3.6 Tahapan Analisis Data.....</b>	<b>35</b>
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>38</b>
<b>4.1 Hasil Penelitian.....</b>	<b>38</b>
<b>4.1.1 Deskripsi Data Pelatihan.....</b>	<b>38</b>
<b>4.1.2 <i>Import Library</i> .....</b>	<b>38</b>
<b>4.1.3 Analisis Statistika Deskriptif.....</b>	<b>39</b>
<b>4.1.4 <i>Data Cleaning</i> .....</b>	<b>41</b>
<b>4.1.5 Pembagian Data <i>Training</i> dan Data <i>Testing</i>.....</b>	<b>42</b>
<b>4.1.6 <i>Random Forest</i> .....</b>	<b>43</b>
<b>4.1.7 <i>Decision Tree</i> .....</b>	<b>43</b>
<b>4.1.8 Perbandingan Metode <i>Random Forest</i> dan <i>Decision Tree</i> .....</b>	<b>44</b>
<b>4.1.9 Prediksi dengan <i>Random Forest</i>.....</b>	<b>45</b>
<b>4.1.10 <i>Variabel Importance</i> .....</b>	<b>46</b>
<b>4.2 Pembahasan .....</b>	<b>46</b>
<b>BAB V KESIMPULAN .....</b>	<b>48</b>
<b>5.1 Kesimpulan .....</b>	<b>48</b>
<b>5.2 Saran.....</b>	<b>49</b>
<b>DAFTAR PUSTAKA .....</b>	<b>50</b>
<b>LAMPIRAN.....</b>	<b>54</b>