

**PEMODELAN KEMISKINAN PROVINSI JAWA TIMUR
MENGUNAKAN METODE CLASSIFICATION AND REGRESSION
TREES DAN BOOTSTRAP AGGREGATING**

SKRIPSI

**OLEH
SAFRILA SHOFANIA
NIM. 16610015**



**JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2020**

**PEMODELAN KEMISKINAN PROVINSI JAWA TIMUR
MENGUNAKAN METODE CLASSIFICATION AND REGRESSION
TREES DAN BOOTSTRAP AGGREGATING**

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S. Mat)**

**Oleh
Safrila Shofania
NIM. 16610015**

**JURUSAN MATEMATIKA
FAKULTAS SAINS DA TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2020**

**PEMODELAN KEMISKINAN PROVINSI JAWA TIMUR
MENGUNAKAN METODE CLASSIFICATION AND REGRESSION
TREES DAN BOOTSTRAP AGGREGATING**

SKRIPSI

Oleh
Safri Shofania
NIM. 16610015

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal 17 Juni 2020

Pembimbing I,

Pembimbing II,

Ria Dhea Layla Nur Karisma, M.Si
NIDT. 19900709201802012228

Muhammad Khudzaifah, M.Si
NIDT. 19900511201608011057

Mengetahui
Ketua Jurusan Matematika

Dr. Usman Pagalay, M.Si
NIP. 196504142003121001

**PEMODELAN KEMISKINAN PROVINSI JAWA TIMUR
MENGUNAKAN METODE CLASSIFICATION AND REGRESSION
TREES DAN BOOTSTRAP AGGREGATING**

SKRIPSI

Oleh
Safri Shofania
NIM. 16610015

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S. Mat)
Tanggal 24 Juni 2020

Penguji Utama : Abdul Aziz, M.Si
Ketua Penguji : Dr. Sri Harini, M.Si
Sekretaris Penguji : Ria Dhea Layla N.K, M.Si
Anggota Penguji : Muhammad Khudzaifah, M.Si

()
()
()
()

Mengetahui
Ketua Jurusan Matematika


Dr. Usman Pagalay, M.Si
NIP. 196504142003121001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Safrila Shofania

NIM : 16610015

Jurusan : Matematika

Fakultas : Sains dan Teknologi

Judul Skripsi : Pemodelan Kemiskinan Provinsi Jawa Timur

Menggunakan Metode *Clasisification And Regression Trees Dan Bootstrap Aggregating*

menyatakan dengan sebenarnya bahwa skripsi yang saya tulis benar-benar merupakan hasil karya saya sendiri, bukan merupakan prngambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar rujukan. Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 14 Juni 2020
Yang membuat Pernyataan



Safrila Shofania
NIM. 16610015

MOTTO

“Rintangan merupakan jembatan menuju kesuksesan”



PERSEMBAHAN

Skripsi ini penulis persembahkan untuk:

Ayahanda Eko Hariadi dan Ibunda Luluk Farida tercinta yang senantiasa dengan ikhlas dan istiqomah mendo'akan, memberi nasihat, dan kasih sayang yang tak ternilai. Seluruh dosen yang senantiasa memberikan ilmu serta bimbingan dengan ikhlas serta sahabat saya Doni yang membantu penyempurnaan format serta selalu memberikan dukungan serta semangat apapun rintangan dalam penyusunan skripsi ini.



KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji bagi Allah Swt atas rahmat, taufik serta hidayah-Nya, sehingga penulis mampu menyelesaikan penyusunan skripsi ini sebagai salah satu syarat memperoleh gelar sarjana dalam bidang matematika di Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Proses penyusunan skripsi ini mendapat banyak bimbingan dan arahan dari berbagai pihak. Sehingga, ucapan terimakasih yang sebesar-besarnya dan penghargaan yang setinggi-tingginya penulis sampaikan terutama kepada:

1. Prof. Dr. H. Abd Haris, M. Ag, selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Sri Harini, M.Si, selaku sekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Usman Pagalay, M.Si, selaku ketua jurusan Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Ria Dhea Layla Nur Karisma, M.Si, selaku dosen pembimbing I yang telah banyak memberikan arahan, nasihat, motivasi kepada penulis sehingga dapat menyelesaikan penyusunan skripsi dengan baik.
5. Muhammad Khudzaifah, M.Si, selaku dosen pembimbing II yang telah memberikan bimbingan serta arahan kepada penulis.
6. Segenap civitas akademika Jurusan Matematika, Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang, terutama untuk seluruh dosen matematika terimakasih atas segala ilmu dan bimbingannya.
7. Segenap keluarga terutama Ayah dan Ibu yang selalu memberikan do'a terbaik bagi penulis.
8. Sahabat-sahabat terbaik penulis, yang selalu menemani, membantu dan memberikan dukungan sehingga penulis dapat menyelesaikan penyusunan skripsi ini.

9. Seluruh pihak yang ikut membantu dalam menyelesaikan skripsi ini baik moril maupun materiil.

Semoga Allah Swt melimpahkan Rahmat dan karunia-Nya kepada kita semua. Penulis berharap semoga dengan rahmat dan izin-Nya mudah-mudahan skripsi ini bermanfaat bagi penulis dan bagi pembaca.

Ammiinn.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Malang, 17 Juni 2020



Penulis



DAFTAR ISI

| | |
|--------------------------------------------------------------------|------|
| HALAMAN JUDUL | |
| HALAMAN PENGAJUAN | |
| HALAMAN PERSETUJUAN | |
| HALAMAN PENGESAHAN | |
| HALAMAN PERNYATAAN KEASLIAN TULISAN | |
| HALAMAN MOTO | |
| HALAMAN PERSEMBAHAN | |
| KATA PENGANTAR | viii |
| DAFTAR ISI | x |
| DAFTAR TABEL | xiii |
| DAFTAR GAMBAR | xiv |
| ABSTRAK | xv |
| ABSTRACT | xvi |
| ملخص | xvii |
| | |
| BAB I PENDAHULUAN | |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah..... | 5 |
| 1.3 Tujuan | 5 |
| 1.4 Batasan Masalah | 5 |
| 1.5 Manfaat | 6 |
| 1.6 Sistematika Penulisan | 6 |
| | |
| BAB II KAJIAN PUSTAKA | |
| 2.1 Classification and Regression Trees (CART)..... | 8 |
| 2.2 Variansi dan Kovariansi..... | 8 |
| 2.3 Metode Pohon Regresi | 9 |
| 2.4 Langkah-Langkah Metode Pohon Regresi..... | 11 |
| 2.4.1 <i>Tree Building</i> (Pembentukan Pohon) | 11 |
| 2.4.2 <i>Pruning Tree</i> (Pemangkasan Pohon)..... | 13 |
| 2.4.3 <i>Optimal Tree Selection</i> (Pemilihan Pohon Optimal)..... | 15 |
| 2.5 Root Mean Squared Error (RMSE)..... | 20 |
| 2.6 Metode Bootstrap..... | 21 |
| 2.7 Bootstrap Aggregating (Bagging)..... | 22 |

| | | |
|-------|----------------------------------------------------------------------|----|
| 2.7.1 | Langkah-langkah Bootstrap Aggregating (bagging) | 25 |
| 2.8 | Kemiskinan menurut Badan Pusat Statistik (BPS) | 26 |
| 2.8.1 | Kemiskinan di Jawa Timur | 26 |
| 2.8.2 | Variabel yang Berpengaruh Terhadap Kemiskinan di Jawa Timur | 27 |
| 2.9 | Integrasi Al-Qur'an | 28 |

BAB III METODOLOGI PENELITIAN

| | | |
|-----|-------------------------------|----|
| 3.1 | Pendekatan Penelitian | 31 |
| 3.2 | Jenis dan Sumber Data | 31 |
| 3.3 | Variabel Penelitian | 31 |
| 3.4 | Metode Analisis Data | 32 |
| 3.5 | Diagram Alur Penelitian | 34 |

BAB IV HASIL DAN PEMBAHASAN

| | | |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 4.1 | Deskriptif Data | 35 |
| 4.1.1 | Deskripsi Kemiskinan Provinsi Jawa Timur | 36 |
| 4.1.2 | Deskripsi Indeks Pembangunan Manusia (IPM) | 37 |
| 4.1.3 | Deskripsi Tingkat Partisipasi Angkatan Kerja (TPAK) | 38 |
| 4.1.4 | Deskripsi Tingkat Pengangguran Terbuka (TPT) | 39 |
| 4.1.5 | Deskripsi Pengeluaran Perkapita (Riil yang Disesuaikan) Provinsi Jawa Timur | 39 |
| 4.1.6 | Deskripsi Wanita Umur 15-40 Pernah Kawin dan Melahirkan Hidup dengan Pertolongan Dokter Provinsi Jawa Timur (PK) | 40 |
| 4.1.7 | Deskripsi Jumlah Rumah Tangga Usaha Pertanian (Jenis Usaha Pertanian Utama yang Diusahakan, 2018) Provinsi Jawa Timur (SP) | 41 |
| 4.1.8 | Deskripsi PDRB (Atas Dasar Harga Konstan 2010) Provinsi Jawa Timur | 41 |
| 4.2 | Pemodelan Kemiskinan Provinsi Jawa Timur Menggunakan Metode CART dan <i>Bagging</i> CART | 42 |
| 4.2.1 | Model Classification And Regression Trees (CART) | 42 |
| 4.2.2 | Model Bootstrap Aggregating | 47 |
| 4.2.3 | Perbandingan Keakuratan Model | 48 |
| 4.3 | Variabel yang mempengaruhi kemiskinan Provinsi Jawa Timur menurut metode <i>Classification And Regression Trees</i> dan <i>Bootstrap Aggregating</i> . 49 | |

BAB V PENUTUP

| | | |
|-----|------------------|----|
| 5.1 | Kesimpulan | 52 |
| 5.2 | Saran | 52 |

DAFTAR PUSTAKA54

LAMPIRAN



DAFTAR TABEL

| | | |
|------------|---------------------------------------------------------------------------------------------------------------------|----|
| Tabel 3.1 | Variabel Respon dan Prediktor Kemiskinan Jawa Timur | 32 |
| Tabel 4.1 | Kemiskinan Provinsi Jawa Timur Tahun 2018..... | 37 |
| Tabel 4.2 | IPM Provinsi Jawa Timur Tahun 2018..... | 37 |
| Tabel 4.3 | TPAK Provinsi Jawa Timur Tahun 2018..... | 38 |
| Tabel 4.4 | TPT Jawa Timur Tahun 2018 | 39 |
| Tabel 4.5 | Pengeluaran Perkapita (Rupiah) Provinsi Jawa Timur tahun 2018 | 40 |
| Tabel 4.6 | Persentase Melahirkan Hidup Provinsi Jawa Timur Tahun 2018..... | 40 |
| Tabel 4.7 | Sektor Pertanian (Individu) Provinsi Jawa Timur Tahun 2018 | 41 |
| Tabel 4.8 | PDRB Provinsi Jawa Timur Tahun 2018..... | 42 |
| Tabel 4.9 | Perbandingan Keakuratan metode <i>Classification And Regression Trees</i> dan <i>Bootstrap Aggregating</i> | 48 |
| Tabel 4.10 | <i>Variable Importance Classification And Regression Trees</i> | 49 |
| Tabel 4.11 | <i>Variable Importance Bootstrap Aggregating</i> | 50 |

DAFTAR GAMBAR

| | |
|--------------------------------------------------------------------------------|----|
| Gambar 2.1 Pohon Regresi T | 13 |
| Gambar 2.2 Cabang Pohon Regresi Tt_2 | 14 |
| Gambar 2.3 Pohon Regresi $T - Tt_2$ | 14 |
| Gambar 2.4 Prosedur Pelatihan <i>Bagging</i> | 24 |
| Gambar 2.5 Prediksi dengan Model <i>Training</i> | 25 |
| Gambar 3.1 Diagram Alur Pemodelan CART dan <i>Bagging</i> CART | 34 |
| Gambar 4.1 Grafik Persentase Kemiskinan di Jawa Timur | 36 |
| Gambar 4.2 Plot <i>Matrix</i> Kemiskinan Provinsi Jawa Timur | 43 |
| Gambar 4.3 <i>Output</i> Pohon Regresi Maksimal | 44 |
| Gambar 4.4 Plot Parameter Kompleksitas Pemangkasan (cp) | 45 |
| Gambar 4.5 Plot Parameter Kompleksitas Pemangkasan (cp) Pohon Dewasa | 46 |
| Gambar 4.6 <i>Output</i> Pohon Regresi Optimal | 47 |

ABSTRAK

Shofania, Safrila. 2020. **Pemodelan Kemiskinan Provinsi Jawa Timur Menggunakan Metode *Classification And Regression Trees* dan *Bootstrap Aggregating***. Skripsi. Jurusan Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Pembimbing: (1) Ria Dhea Layla Nur Kharisma, M.Si (2) Muhammad Khudzaifah, M.Si

Kata Kunci: Kemiskinan, *Classification And Regression Trees*, *Bootstrap*, *Bootstrap Aggregating*

Kemiskinan merupakan salah satu problematika dalam perekonomian Indonesia yang terbilang kompleks dan mendasar. Upaya pencarian solusi untuk mengatasi sebuah kemiskinan menjadi hal yang rumit serta menjadi salah satu program prioritas dari pemerintah dimana penduduk miskin yang besar akan menjadi beban bagi kemajuan sebuah negara. Provinsi Jawa Timur memiliki tingkat kemiskinan yang cukup tinggi, dimana pada tahun 2018 jumlah penduduk miskin menurut BPS Provinsi Jawa Timur adalah sebanyak 4332,59 jiwa dimana telah mengalami penurunan dari tahun sebelumnya, akan tetapi masih diperlukan solusi untuk mengetahui variabel apa yang mempengaruhi untuk terjadinya kemiskinan tersebut. Tujuan dari penelitian ini adalah untuk mengetahui pemodelan kemiskinan di Provinsi Jawa Timur berdasarkan variabel yang sangat berpengaruh dalam terbentuknya sebuah kemiskinan di sebuah daerah dengan menggunakan metode statistik nonparametrik *Classification And Regression Trees* dan *Bootstrap Aggregating*. Hasil dari penelitian menunjukkan bahwa metode yang dimiliki oleh metode *Bootstrap Aggregating* lebih tinggi dari pada keakuratan dari metode *Classification And Regression Trees*. Metode *Bootstrap Aggregating* memperoleh nilai *RMSE* sebesar 2,881789 dan nilai *RMSE* yang diperoleh dari metode *Classification And Regression Trees* adalah sebesar 3,498211. Variabel yang mempengaruhi terjadinya kemiskinan di Provinsi Jawa Timur pada tahun 2018 menurut metode *Classification And Regression Trees* pada urutan pertama yaitu variabel PDRB dan urutan pertama pada metode *Bootstrap Aggregating* adalah variabel IPM.

ABSTRACT

Shofania, Safrila. 2020. **Modeling of Poverty in East Java Province Using Classification And Regression Trees And Bootstrap Aggregating**. Thesis. Mathematics Department, Science and Technology Faculty, State Islamic University Of Maulana Malik Ibrahim Malang. Advisors: (1) Ria Dhea Layla Nur Kharisma, M.Si (2) Muhammad Khudzaifah, M.Si

Keywords: Poverty, Classification And Regression Trees, Bootstrap, Bootstrap Aggregating.

Poverty is one of the problems in the Indonesian economy which is fairly complex and fundamental. The efforts to find solutions to alleviate poverty are complicated and one of the priority programs of the government where a large number of poor people will become a burden for the progress of a country. East Java Province has a fairly high poverty rate, wherein 2018 the number of poor people according to the BPS of East Java Province was 4332.59 people which has decreased from the previous year, but solutions are still needed to find out what variables affect poverty. The purpose of this study was to determine poverty modeling in East Java Province based on variables that are very influential in the formation of poverty in an area using non-parametric statistical methods Classification And Regression Trees and Bootstrap Aggregating. The results of the study indicate that the method owned by the Bootstrap Aggregating method is higher than the accuracy of the Classification and Regression Trees method. The Bootstrap Aggregating method obtains a RMSE value of 2,881789 and the RMSE value obtained from the Classification and Regression Trees method is equal to 3,498211. The variables that affect the occurrence of poverty in East Java Province in 2018 according to the Classification and Regression Trees method in the first place, namely the PDRB variable and the first order in the Bootstrap Aggregating method is the IPM variable.

ملخص

سوفانيا ، سفريلا. ٢٠٢٠. تصميم الفقر على ولاية جاوى الشرقية بالطريقة (CART) و (Bagging CART). البحث الجامعي. قسم الرياضيات. كلية العلوم والتكنولوجيا جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج.

المشرفة : (١) ريا ديا ليل نور حارما الماجستير، (٢) محمد حظيفة الماجستير
الكلمات الأساسية : الفقر،

من إحدى المشكلات العامة التي تحدث في بلادنا إندونيسيا هي الفقر. والحلول لتلك المشكلة هو من وظيفة بلاد إندونيسيا، لأن تلك المشكلة تكون مانعة و نهضة لتقدم البلاد. ويفوق الفقر في ولاية جاوى الشرقية لأن المساكين فيها في سنة ٢٠١٨ عند BPS ٤٣٣٢.٥٩ التي تنزل من قبلها. لكن يطلب سبب ذلك. والهدف من هذه البحث ليعلم مثال المسكين فيها بطريقة Classification And Regression Trees و Bootstrap Aggregating. و نتائج البحث هو أن الطريقة المملوكة لطريقة Bootstrap Aggregating أدق من طريقة Classification And Regression Trees. حصلت طريقة تجميع Bootstrap على قيمة RMSE تبلغ ٢.٨٨١٧٨٩ وكانت قيمة RMSE التي تم الحصول عليها من Classification And Regression Trees ٣.٤٩٨٢١١. والمتقلب المؤثر من الفقر في جاوى الشرقية سنة ٢٠١٨ عند طريقة Classification And Regression Trees في دربة الأولى متقلب PDRB و عند Bootstrap Aggregating هي متقلب IPM.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Negara berkembang seperti Indonesia pada umumnya memiliki permasalahan utama dalam pembangunan ekonomi yang disebabkan oleh pendapatan yang rendah. Pertumbuhan ekonomi merupakan salah satu indikator dalam mengentaskan kemiskinan dimana pertumbuhan dari ekonomi adalah suatu konsep dari pembangunan ekonomi. Sebagian besar penduduk di Indonesia masih banyak ditemukan masyarakat dengan kualitas hidup yang rendah. Ledakan jumlah penduduk yang bertambah setiap tahunnya juga menjadi salah satu faktor pemicu kemiskinan dengan memberikan dampak buruk bagi kehidupan sosial ekonomi masyarakat yaitu jumlah pengangguran yang semakin meningkat serta terjadinya kekurangan pangan dan angka kemiskinan semakin melonjak (Qurratu'ain, 2016:265).

Badan Pusat Statistik (BPS:2019) menunjukkan laju kemiskinan di provinsi Jawa Timur pada tahun 2011 hingga tahun 2018 selalu mengalami penurunan, persentase kemiskinan pada tahun 2017 sebesar 11,77% dan mengalami penurunan pada tahun 2018 yaitu sebesar 10,98% dengan jumlah 4332,59 penduduk miskin di Jawa Timur. Tingkat kemiskinan di Jawa Timur merupakan tingkat kemiskinan dari 38 Kabupaten/ Kota di Jawa Timur. Jumlah penduduk miskin yang tidak merata di Jawa Timur mencerminkan belum adanya tingkat pemerataan pendapatan di masyarakat walaupun pertumbuhan ekonomi Jawa Timur cukup tinggi. Sektor pertanian memiliki tingkat produktivitas tenaga kerja paling rendah dibandingkan dengan sektor-sektor lainnya, oleh karena itu memerlukan sebuah perubahan paradigma indikator pembangunan ekonomi pertumbuhan ekonomi yang hanya menghitung produk domestik bruto (PDB) menjadi paradigma pertumbuhan ekonomi yang menambahkan indikator lain seperti halnya indikator pemerataan pendapatan (Azizah, 2018:168).

Persentase kemiskinan di Jawa timur menurut BPS (2019) dari tahun ke tahun semakin menurun, akan tetapi penurunan persentase tidak secara langsung dalam jumlah yang banyak, penurunan persentase kemiskinan yang dialami

kurang lebih sebanyak 1,00% per tahun. Penurunan persentase kemiskinan disebabkan oleh banyak faktor, sehingga pengklasifikasian menggunakan pohon keputusan digunakan untuk mengetahui faktor apa saja yang mendominasi terjadinya kemiskinan. Pengklasifikasian masalah kemiskinan dapat dilakukan menggunakan metode *Classification and Regression Trees* dan *Bootstrap Aggregating* yang merupakan analisis non-parametrik yang bertujuan untuk menyajikan kaidah keputusan berbentuk pohon biner yang membagi data pada *learning* sampel dengan batasan linier univariat. Analisis *Classification and Regression Trees* dimulai dari *node root* dan berakhir pada pengamatan yang memiliki nilai 0.

Pengklasifikasian *Classification and Regression Trees* akan menghasilkan sebuah pohon regresi apabila variabel yang digunakan bersifat kontinu (Ratnaningrum,2016: 82). Pohon regresi (*regression trees*) merupakan regresi berstruktur pohon yang menggunakan kaidah keputusan (*decision trees*) yang dibentuk melalui algoritma penyekatan yang dilakukan secara bertahap. Pohon regresi memiliki tujuan utama diantaranya adalah untuk menghasilkan sebuah pohon prediktor atau kaidah prediksi (Yohanes,1999:21). Prediktor dibangun dengan mendeteksi heterogenitas yang ada pada data dan terdapat dua kaidah pemisah yaitu *Least Square* (LS) dan *Function the Least Absolute Deviation* (LAD).

Berbagai penelitian telah dilakukan untuk menyelesaikan masalah yang berkaitan dengan pengklasifikasian khususnya menggunakan *Regression Trees*. Penelitian tersebut dilakukan oleh Liqun Diao dan Chengguo Weng (2019) dengan judul penelitian '*Regression Tree Credibility Model*' yang menerapkan metode *regression trees* pada model kredibilitas (*credibility Model*) dimana *Regression trees* merupakan pengembangan dari model prediksi yang efektif. Peneliti selanjutnya dengan judul '*Toward A Simplification Of Models Using Regression Trees*' karya Y. Eynaud, D. Nerini, M dan J.-C Poggiale (2019) yang membahas tentang penyederhanaan model menggunakan *Regression Trees* yang diketahui bahwa optimasi parameter yang digunakan menjadi lebih mudah dan mungkin bisa menjadi lebih kuat. Penelitian lainnya juga dilakukan oleh Wieta B Komalasari dalam kasus survei pendapatan tani 2014 untuk mengetahui faktor apa

saja yang mempengaruhi pendapatan usaha tani dengan judul penelitiannya yaitu ‘Metode Pohon Regresi Untuk Eksploratori Data dengan Peubah yang Banyak dan Kompleks’.

Decision tree biasanya telah dikombinasikan dengan pendekatan *Bootstrap Aggregating (bagging)* untuk membentuk estimator yang lebih kuat, yang merupakan proses umum untuk mengurangi nilai variansi CART yang tinggi (Xie, 2019:2). Metode *Bootstrap Aggregating* telah banyak diterapkan dalam bidang-bidang seperti biostatistik dan *remote sensing*, akan tetapi metode *Bootstrap Aggregating* penggunaannya tidak biasa dalam bidang ekologi (Prasad, 2006:184). Penelitian sebelumnya mengenai metode *Bootstrap Aggregating* pada pohon regresi pernah dilakukan oleh Jiemin Xie (2019) dengan penelitiannya yang berjudul ‘*Ensemble Of Bagged Regression Tree For Concrete Dam Deformation Predicting*’ yang membahas tentang ensemble *Bootstrap Aggregating* pohon regresi untuk memprediksi deformasi dari bendungan beton, yang memperoleh hasil apabila dengan menggunakan metode *Bootstrap Aggregating* model yang dihasilkan memiliki efek prediksi yang baik dan kinerja metodenya lebih baik dari pada metode pohon regresi. Metode *Classification And Regression Trees* dan *Bootstrap Aggregating* pada penelitian ini akan dibandingkan yang akan diukur dengan menggunakan dengan nilai *RMSE* yang menurut Alboukadel Kassambara (2017) pada bukunya yang berjudul ‘*Machine Learning Essential’s* menjelaskan apabila *RMSE* merupakan rata-rata dari nilai yang diperoleh dari hasil pengamatan dan nilai prediksi model.

Berdasarkan pemaparan diatas, penelitian ini difokuskan pada tingkat kemiskinan Provinsi Jawa Timur pada tahun 2018 menggunakan metode *Classification And Regression Trees* dan *Bootstrap Aggregating* Ayat Al-Qur’an yang berhubungan dengan masalah kemiskinan salah satunya yaitu pada Q.S Al-Baqarah ayat ke-61 dimana Al-Qur’an didalamnya mengandung 33 ayat yang berkaitan dengan kata miskin, yang sebagian besar berasal dari kata dasar as-sakan sebanyak 27 ayat (Cahya, 2015:45).

وَإِذْ قُلْتُمْ يُمُوسَىٰ لَنْ نَّصْبِرَ عَلَىٰ طَعَامٍ وَاجِدِ فَادْعُ لَنَا رَبَّكَ يُخْرِجْ لَنَا مِمَّا تُنْبِثُ الْأَرْضُ مِنْ بَقْلِهَا وَقِثَابِهَا
وَفُؤْمِهَا وَعَدَسِيهَا وَبَصِلِهَا قَالَ آتَسْتَبْدِلُونَ الَّذِي هُوَ أَدْنَىٰ بِالَّذِي هُوَ خَيْرٌ إِهْبَطُوا مِصْرًا فَإِنَّ لَكُمْ مَّا سَأَلْتُمْ

وَضُرِبَتْ عَلَيْهِمُ الذِّلَّةُ وَالْمَسْكَنَةُ وَبَاءُوا بِغَضَبٍ مِّنَ اللَّهِ ذَلِكَ بِأَنَّهُمْ كَانُوا يَكْفُرُونَ بِآيَاتِ اللَّهِ وَيَقْتُلُونَ النَّبِيَّاتِ بِغَيْرِ الْحَقِّ ذَلِكَ بِمَا عَصَوْا وَكَانُوا يَعْتَدُونَ

Q.S Al-Baqarah tersebut memiliki arti “Dan (Ingatlah) ketika kamu berkata: ‘Hai Musa, kami tidak akan sabar (tahan) dengan satu macam makanan saja. Sebab itu mohonkanlah untuk kami kepada Tuhanmu, agar dia mengeluarkan bagi kami apa yang di tumbuhkan bumi, yaitu sayur mayurnya, ketimunnya, bawang putihnya, kacang adasnya dan bawang merahnya’. Musa berkata ‘ maukah kamu mengambil yang rendah sebagai pengganti yang lebih baik? Pergilah kamu ke suatu kota, pasti kamu memperoleh apa yang kamu minta’. Lalu ditimpahkanlah kepada mereka nista dan kehinaan serta mereka mendapat kemurkaan dari Allah. Hal itu (terjadi) karena mereka selalu mengingkari ayat-ayat Allah dan membunuh para nabi yang memang tidak dibenarkan. Demikian itu (terjadi) karena mereka selalu berbuat durhaka dan melampaui batas”.

Keterkaitan ayat tersebut dengan kemiskinan adalah mereka (rakyat) senantiasa selalu mengikuti hawa nafsu yang terus menerus menolak apa yang telah dipikirkan oleh Allah. Akibat dari perbuatan yang telah dilakukan adalah mereka harus menerima kehinaan, kemiskinan serta kesengsaraan. Allah telah murka karena mereka telah berbuat diluar batas.

Berdasarkan uraian mengenai kemiskinan serta metode yang tepat untuk mengetahui faktor yang mendominasi terjadinya kemiskinan, penelitian ini fokus pada pemodelan kemiskinan Provinsi Jawa Timur menggunakan pendekatan metode statistika yaitu *Classification and Regression Trees* dan *Bootstrap Aggregating* yang diaplikasikan pada data faktor-faktor kemiskinan di Jawa Timur tahun 2018. Metode yang digunakan dalam penelitian ini yaitu pendekatan metode *Classification and Regression Trees* yang memiliki beberapa kelebihan yaitu hasil yang diperoleh mudah untuk diinterpretasikan dan lebih cepat perhitungannya. Metode *Classification and Regression Trees* pada penelitian ini akan dilengkapi dengan metode *Bootstrap Aggregating* untuk memperbaiki stabilitas kekuatan prediksi dikarenakan metode *Classification and Regression Trees* memiliki kekurangan yaitu pohon yang dihasilkan cenderung tidak stabil dimana terjadi perubahan pada data *learning* yang berpengaruh terhadap hasil akurasi prediksi. Berdasarkan uraian latar belakang, pada penelitian ini penulis memilih judul “Pemodelan Kemiskinan Provinsi Jawa Timur Menggunakan Metode *Classification and Regression Trees* Dan *Bootstrap Aggregating*”.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka disusun perumusan masalah sebagai berikut:

1. Bagaimana keakuratan dari model yang dihasilkan pada data kemiskinan di Provinsi Jawa Timur tahun 2018 dengan menggunakan metode *Clasification and Regression Trees* dan *Bootstrap Aggregating*?
2. Variabel apa saja yang sangat mempengaruhi terjadinya kemiskinan di Provinsi Jawa Timur tahun 2018 menggunakan metode *Clasification and Regression Trees* dan *Bootstrap Aggregating*?

1.3 Tujuan

Berdasarkan rumusan masalah penelitian tersebut, maka tujuan penelitian dari skripsi ini adalah:

1. Mengetahui keakuratan dari model yang dihasilkan pada data kemiskinan di Provinsi Jawa Timur tahun 2018 dengan menggunakan metode *Clasification and Regression Trees* dan *Bootstrap Aggregating*.
2. Mengetahui variabel apa saja yang sangat mempengaruhi terjadinya kemiskinan di Provinsi Jawa Timur tahun 2018 menggunakan metode *Clasification and Regression Trees* dan *Bootstrap Aggregating*.

1.4 Batasan Masalah

Berdasarkan uraian latar belakang yang telah dipaparkan, berikut ini adalah batasan masalah yang digunakan pada penelitian ini:

1. Data yang digunakan merupakan data statistik dasar Badan Pusat Statistik (BPS) Provinsi Jawa Timur tahun 2018.
2. Variabel yang digunakan sebanyak 8 variabel, variabel-variabel faktor kemiskinan Provinsi Jawa Timur yang digunakan adalah:
 - a. Persentase kemiskinan (%)
 - b. Indeks Pembangunan Manusia (IPM) (%)
 - c. Tingkat Partisipasi Angkatan Kerja (TPAK) (%)
 - d. Tingkat Pengangguran Terbuka (TPT) (%)
 - e. Rumah tangga usaha pertanian (SP) (orang)

- f. Pengeluaran perkapita (PP) (Rupiah)
 - g. Persentase Wanita umur 15-40 pernah kawin dan melahirkan dengan pertolongan dokter (PK) (%)
 - h. Produk Domestik Regional Bruto (PDRB) (%)
3. Metode yang digunakan pada penelitian ini adalah *Classification and Regression Tress* dan *Bootstrap Aggregating*.

1.5 Manfaat

Berdasarkan rumusan masalah dan tujuan penelitian, maka manfaat dari penulisan penelitian ini adalah:

1. Manfaat yang dapat diperoleh dari penelitian ini yaitu menambah wawasan mengenai analisis tentang *Classification and Regression Trees* dan *Bootstrap Aggregating*.
2. Membantu meningkatkan keetelitian pada kasus penanganan kemiskinan yang ada di daerah Jawa Timur sehingga mengetahui faktor penyebab kemiskinan tertinggi.
3. Mengetahui program yang digunakan untuk mengetahui penyebab kemiskinan pada Provinsi Jawa Timur pada tahun 2018 dengan menggunakan metode *Classification and Regression Trees* dan *Bootstrap Aggregating*.

1.6 Sistematika Penulisan

Adapun sistematika penulisan ini terdiri dari empat bab yang terdiri dari beberapa sub bab dengan susunan penulisan sebagai berikut:

BAB I PENDAHULUAN

Bagian pendahuluan merupakan gambaran secara umum diambilnya kasus serta pengangkatan judul yang digunakan diulas didalam latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bagian tinjauan pustaka membahas tentang gambaran yang telah diangkat dan dijelaskan secara umum pada latar belakang dan kembali

dijelaskan lebih mengerucut pada bagian ini. Teori ini diambil dari buku, *literature*, jurnal serta artikel-artikel di internet.

BAB III METODOLOGI PENELITIAN

Bagian metodologi penelitian menjelaskan tentang jenis penelitian yang diangkat, waktu serta tempat penelitian, disertakan juga jenis dan sumber data dari kasus penelitian, teknik pengumpulan data, variabel dan definisi operasional dan juga prosedur penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bagian hasil dan pembahasan menjelaskan tentang hasil yang telah diolah serta pembahasan dari proses penelitian yang dilakukan.

BAB V KESIMPULAN DAN SARAN

Bagian kesimpulan dan saran menjelaskan mengenai hasil akhir/ kesimpulan serta saran dari penelitian yang telah dilakukan.



BAB II KAJIAN PUSTAKA

2.1 Classification and Regression Trees (CART)

Classification and Regression Trees merupakan salah satu metode non-parametrik yang digunakan untuk menganalisis dengan teknik pohon keputusan, baik untuk variabel kategorik maupun kontinu. Variabel respon berskala kontinu akan menghasilkan sebuah pohon regresi (*regression trees*) dan apabila variabel berskala kategorik maka pohon yang dihasilkan adalah pohon klasifikasi (*classification trees*). Metode *Classification and Regression Trees* bertujuan untuk mendapatkan suatu kelompok data yang akurat sebagai suatu pencari dari suatu pengklasifikasian dan menentukan prediksi struktur data. Metode *Classification and Regression Trees* dapat diterapkan pada data dengan jumlah besar dan variabel yang banyak, yang lebih cepat perhitungannya serta hasil yang lebih mudah untuk diinterpretasikan (Ratnaningrum, 2016: 82).

2.2 Variansi dan Kovariansi

Varian merupakan ukuran korelasi antara dua variabel acak yang sama dengan rata-rata dari masing-masing peubah acak mungkin sama meskipun distribusinya tidak sama. Kovariansi merupakan ukuran korelasi dari dua atau lebih variabel acak, dimisalkan dengan X dan Y adalah variabel random berdistribusi peluang gabungan $f(x, y)$. Kovariansi dari X dan Y apabila diskrit adalah:

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \sum_x \sum_Y (x - \mu_X)(y - \mu_Y)f(x, y) \quad (2.1)$$

Apabila X dan Y kontinu maka:

$$\begin{aligned} \sigma_{XY} &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y) dx dy \end{aligned} \quad (2.2)$$

Variansi memiliki beberapa sifat diantaranya adalah:

1) Teorema 1

Rinaldi (2010:20) menjelaskan jika a dan b adalah suatu konstanta, maka:

$$\sigma^2_{aX+b} = a^2\sigma^2 \quad (2.3)$$

Jika $a = 1$, maka $\sigma^2_{X+b} = \sigma^2_X = \sigma^2$

Jika $b = 0$, maka $\sigma^2_{aX} = a^2\sigma^2_X = a^2\sigma^2$

2) Teorema 2

Rinaldi (2010:21) menjelaskan jika X dan Y merupakan peubah acak berdistribusi peluang $f(x, y)$, maka:

$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY} \quad (2.4)$$

Akibat:

Jika X dan Y peubah acak yang saling bebas, maka:

$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$$

Jika X dan Y variabel random yang saling bebas, maka:

$$\sigma^2_{aX-bY} = a^2\sigma^2_X + b^2\sigma^2_Y$$

2.3 Metode Pohon Regresi

Breiman (1998: 221) menjelaskan bahwa pada pohon regresi terdapat suatu objek yang terdiri dari data (x, y) dimana x merupakan nilai yang berada pada ruang pengukuran X dan y merupakan nilai *real* yang berada pada ruang pengukuran Y . Variabel y disebut dengan variabel respon sedangkan x adalah variabel prediktor. Aturan prediksi adalah fungsi $d(x)$ didefinisikan pada ruang pengukuran X yang menghasilkan sebuah nilai *real*, sehingga diketahui bahwa $d(x)$ bernilai *real*.

Analisis pohon regresi merupakan istilah umum dalam pembentukan prediktor $d(x)$ yang berasal dari sampel *learning* \mathcal{L} . Pembangunan sebuah prediktor atau penduga respon menurut Breiman (1998:221) memiliki dua tujuan yaitu, untuk memprediksi secara tepat serta teliti untuk vektor pengukuran dimasa yang akan datang dengan seakurat mungkin dan yang kedua adalah untuk mengetahui hubungan struktural antara variabel respon dengan variabel prediktor. Sebuah sampel *learning* yang terdiri dari N objek $(x_1, y_1), \dots, (x_n, y_n)$ berguna untuk pembentukan penduga respon $d(x)$ serta $(x'_1, y'_1), \dots, (x'_{N_2}, y^1)$ pada N_2

objek, sehingga keakuratan $d(x)$ dapat diukur menggunakan *mean squared error*, yakni:

$$\frac{1}{N_2} \sum_{n=1}^{N_2} |y'_n - d(x'_n)| \quad (2.5)$$

dengan $d(x'_n)$ merupakan prediktor dari y'_n dan $n = 1, 2, \dots, N_2$.

Breiman (1998:222) juga menjelaskan bila *mean squared error* $R^*(d)$ dari penduga respon didefinisikan sebagai berikut:

$$R^*(d) = \sum (Y - d(X))^2 \quad (2.6)$$

dengan:

$R^*(d)$ = Harapan kesalahan menggunakan prediktor $d(x)$ sebagai penduga respon Y .

Average squared error (rata-rata kesalahan) relatif $RE^*(d)$ juga didefinisikan sebagai berikut:

$$RE^*(d) = \frac{R^*(d)}{R^*(\mu)} \quad (2.7)$$

dengan:

$\mu = E(Y)$ merupakan penduga respon untuk Y apabila nilai X tidak diketahui.

$R^*(\mu) = E(Y - \mu)^2$ merupakan variansi dari Y atau $Var Y$ dan juga merupakan *average square error* (rata-rata kuadrat kesalahan) menggunakan nilai μ .

Breiman (1998:224) juga menuliskan bahwa jika rata-rata serta variansi data uji masing-masing adalah:

$$\text{rata - rata } (\bar{y}) = \frac{1}{N} \sum_n y_n \quad (2.8)$$

$$\text{Variansi } (R(\bar{y})) = \frac{1}{N} \sum_n (y_n - \bar{y})^2 \quad (2.9)$$

maka estimasi pengganti $RE(d)$ menjadi $RE^*(d)$ ialah:

$$\frac{R(d)}{R(y)} \quad (2.10)$$

dengan:

$R(d)$ = estimasi pengganti $R^*(d)$

$R(y)$ = Average square error (rata-rata kesalahan kuadrat) data uji.

Sehingga diketahui rumus dari kesalahan (*error*) relatif dari data uji untuk estimasi $RE^*(d)$ adalah:

$$RE^{ts}(d) = \frac{R^{ts}(d)}{R(\bar{y})}, \quad (2.11)$$

dan rumus kesalahan estimasi uji untuk $R^*(d)$ adalah:

$$R^{ts}(d) = \frac{1}{N_2} \sum_{n=1}^{N_2} (y_n - d(x_n))^2 \quad (2.12)$$

2.4 Langkah-Langkah Metode Pohon Regresi

Roger J Lewis dalam bukunya 'An Introduction to Classification and Regression Trees' (2000:6) menjelaskan bahwa langkah-langkah metode *classification and regression trees* (CART) terjadi dalam beberapa tahap. Tahap pertama adalah *tree building* (pembangunan pohon), *stopping tree building* (penghentian pembangunan pohon), *pruning tree* (pemangkasan pohon), dan *optimal tree selection* (pemilihan pohon optimal). Berikut merupakan penjelasan dari tahap-tahap yang harus dilakukan dalam metode pohon regresi:

2.4.1 Tree Building (Pembentukan Pohon)

Tahap pertama metode pohon regresi adalah pembentukan pohon, dalam pembentukan pohon terdapat proses pemilahan simpul induk menjadi dua simpul anak melalui aturan pemilahan simpul yang dilakukan berulang kali. Pohon regresi dibentuk melalui penyekatan gugus data dengan sederet penyekatan biner sampai dihasilkan simpul akhir (Breiman, 1998 dalam Yuliani, 2014:20). Penyekatan suatu simpul untuk menjadi dua simpul anak dilakukan dengan aturan pemilihan seperti berikut:

- a. Proses penyekatan bergantung pada nilai yang bersal dari satu peubah penjelas.
- b. Peubah kontinu x_j dengan j merupakan variabel prediktor, $j = 1, 2, \dots, N$ menggunakan penyekatan atau pemilahan $x_j \leq c$, dengan c merupakan nilai tengah antara dua nilai amatan peubah x_j secara berurutan. Banyak n nilai yang berbeda yang dimiliki oleh x_j maka akan terdapat $n - 1$ pemilahan.

- c. Variabel kategorik penyekatan atau pemilahan berasal dari semua kemungkinan penyekatan atau pemilahan berdasarkan terbentuknya dua anak gugus yang saling lepas (*disjoint*). Peubah kategorik nominal dengan \mathcal{L} kategori pada x_j , maka akan terjadi $2^{\mathcal{L}-1} - 1$ pemilahan, apabila x_j kategorik ordinal maka penyekatan atau pemilahan yang terjadi adalah $\mathcal{L} - 1$ pemilahan.

Jumlah dari kuadrat deviasi digunakan untuk kriteria kehomogenan pada tiap-tiap simpul t . Jumlah kuadrat deviasi dari simpul t dirumuskan dengan $R(t) = \sum_{x_n \in t} [y_n - \bar{y}(t)]^2$ dengan $\bar{y}(t) = \frac{1}{N(t)} \sum_{x_n \in t} y_n$ dimana y_n merupakan nilai individu variabel respon simpul t , nilai dari x_n merupakan nilai individu variabel prediktor simpul t , $\bar{y}(t)$ merupakan nilai rata-rata semua data y_n pada simpul t , serta $N(t)$ adalah jumlah data dalam simpul t (Breiman, 1998:231).

Breiman (1998:231) mendefinisikan jika pemilah terbaik s^* yang berasal dari simpul t pada S yang memaksimumkan jumlah kuadrat pohon regresi $R(T)$. Lebih tepatnya setiap pemilah s dari t menjadi t_L dan t_R , sehingga persamaannya adalah:

$$\Delta R(s, t) = R(t) - R(t_L) - R(t_R) \quad (2.13)$$

Pemilah terbaik s^* menjadi pemilah terbaik yang menyebabkan:

$$\Delta R(s^*, t) = \max_{s \in S} \Delta R(s, t) \quad (2.14)$$

Persamaan (2.14) merupakan selisih jumlah kuadrat deviasi terbesar yang dijadikan pemilah terbaik dengan S merupakan kumpulan pemilah biner, t_L dan t_R adalah simpul anak kiri dan simpul anak kanan.

2.4.1 *Stopping Tree Building* (Penghentian Pembangunan pohon)

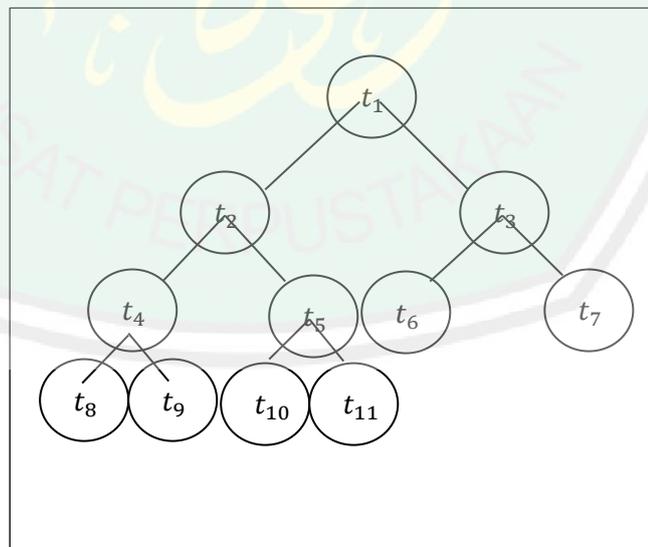
Penghentian pembangunan pohon dilakukan apabila pada tahap pembangunan pohon telah dihasilkan pemilahan atau penyekatan yang terbaik yang kemudian dipilah atau disekat menjadi dua bagian yaitu simpul anak kiri dan kanan (t_L dan t_R). Proses pemilahan atau penyekatan terhadap simpul anak kiri dan kanan (t_L dan t_R) dilakukan berulang kali hingga tak bisa disekat lagi atau dihentikan. Dewi (2012:4) menjelaskan bahwa proses rekursif akan berakhir apabila banyaknya amatan pada simpul akhir ≤ 5 . Kriteria penghentian pembangunan pohon regresi yaitu dengan memaksimumkan kehomogenan ragam

yang terdapat pada setiap simpul dan juga dipengaruhi oleh peubah yang berpengaruh terhadap respon. Simpul-simpul yang tidak lagi mengalami pemilahan akan menjadi simpul terminal atau biasa disebut dengan simpul akhir, sehingga pohon regresi yang terbentuk pada tahap ini dinamakan pohon maksimal (T_{max}).

2.4.2 Pruning Tree (Pemangkasan Pohon)

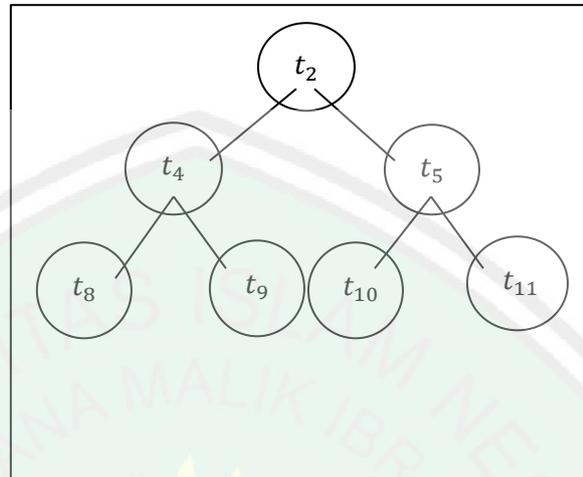
Pohon yang telah melalui tahap pemilahan atau penyekatan serta pembangunan pohon masih menghasilkan pohon regresi yang sangat besar. Tahap penghentian dilakukan hanya karena banyaknya data yang terdapat di simpul akhir atau besarnya peningkatan kehomogenan. Tahap pemangkasan dilakukan untuk memperoleh sebuah pohon akhir yang sederhana dengan memangkas kelebihan dari pohon regresi. Memangkas kelebihan pohon berartikan memangkas pohon (T_{max}) menjadi beberapa pohon regresi yang ukurannya lebih kecil atau sering disebut subpohon.

Simpul t' biasanya disebut dengan anak atau keturunan dari simpul t , dan simpul t sendiri disebut dengan induk atau leluhur dari simpul t' . Berikut penjelasan dari sebuah pohon regresi:



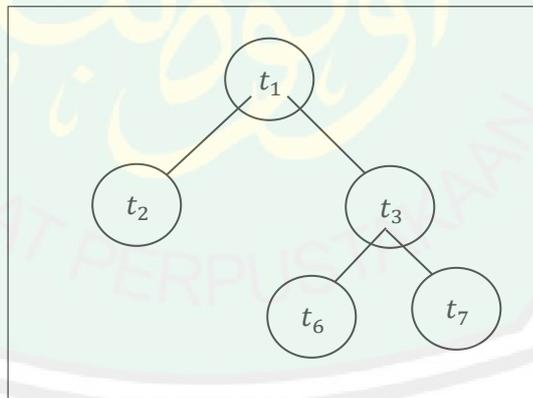
Gambar 2.1 Pohon Regresi T

Gambar 2.1 menjelaskan bahwa pohon regresi T dimana $t_4, t_5, t_8, t_9, t_{10}, t_{11}$ merupakan anak dari t_2 . Simpul t_4, t_2, t_1 merupakan induk atau ancestor dari simpul t_9 .



Gambar 2.2 Cabang Pohon Regresi T_{t_2}

Pohon regresi T_{t_2} terdiri dari simpul pohon t_2 dan simpul anak $t_4, t_5, t_8, t_9, t_{10}, t_{11}$ dan cabang T_t dari T dengan simpul akar $t \in T$ berisikan simpul t itu sendiri dengan semua simpul anak (keturunan) dari t dalam T .



Gambar 2.3 Pohon Regresi $T - T_{t_2}$

Breiman (1998:235) menjelaskan bahwa pemangkasan dari suatu pohon cabang T_t akan menghasilkan simpul anak (keturunan) t dari T dengan pengecualian simpul akarnya (t itu sendiri), penjelasan tersebut telah diaplikasikan pada gambar diatas dimana pemangkasan terjadi pada simpul t_2

maka simpul anak dari t_2 adalah $t_4, t_5, t_8, t_9, t_{10}, t_{11}$ yang dipangkas kecuali simpul akar yaitu simpul t_2 sendiri.

Tahap pemangkasan dimulai dengan mengambil simpul anak kanan (t_R) dan simpul anak kiri (t_L) dari T_{max} yang dihasilkan dari simpul induk t . Apabila persamaan $\Delta R(s, t) = R(t) - R(t_L) - R(t_R)$ menghasilkan nilai 0, maka simpul anak kiri (t_L) dan kanan (t_R) akan dipangkas yang menghasilkan seperti halnya pada persamaan $\Delta R(s^*, t) = \max_{s \in S} \Delta R(s, t)$ atau dapat dituliskan dengan $R(T_1) = R(T_{max})$. Suharjo (2008:42) menjelaskan bahwa inti dari tahap pemangkasan itu sendiri adalah pemotongan hubungan terlemah (*weakest-link cutting*) pada pohon regresi T .

Tahap pemangkasan dilakukan berulang kali sampai tidak terdapat lagi pemangkasan yang mungkin dan diperoleh urutan $T_1, > T_2 > \dots > T_n$ dimana $T_n = \{t_1\}$ sehingga menjadi $T_1, > T_2 > \dots > \{t_1\}$. Besar rata-rata kuadrat kesalahan didefinisikan sebagai berikut:

$$R(T_t) = \sum_{t \in T_t} R(t') \quad (2.15)$$

dengan:

$R(T_t)$ = kuadrat tengah galat subpohon

$\sum_{t \in T_t} R(t')$ = jumlah kuadrat tengah simpul kiri dan simpul kanan hasil dari pemilahan atau penyekatan.

2.4.3 Optimal Tree Selection (Pemilihan Pohon Optimal)

Pohon regresi yang berukuran besar dan kompleks yang terbentuk dapat menggambarkan sebuah struktur data, sehingga *pruning tree* (pemangkasan pohon) perlu dilakukan. Tahap pemangkasan terjadi dengan membangun pohon menggunakan *data learning* selanjutnya menggunakan pohon yang terbentuk untuk *data testing*. Breiman (1998:234) menjelaskan bahwa untuk menduga tingkat kesalahan prediksi pohon regresi dengan menggunakan nilai kesalahan estimasi *data testing*. *Data learning* \mathcal{L}_1 dan *data testing* \mathcal{L}_2 merupakan data pengamatan yang telah diacak, dimana *data learning* \mathcal{L}_1 digunakan untuk membuat urutan $\{T_k\}$ dari *pruning tree* (pemangkasan pohon). Berikut merupakan kesalahan estimasi *data testing*:

$$R^{ts}(T_k) = \frac{1}{N_2} \sum_{(x_n, y_n) \in \mathcal{L}_2} (y_n - d_k(x_n))^2 \quad (2.16)$$

dengan:

N_2 = jumlah *data testing*

x_n = nilai individu variabel prediktor yang ada pada *data testing*

y_n = nilai individu variabel respon yang ada pada *data testing*

$d_k(x_n)$ = dugaan respon dari pengamatan ke n^* pada pohon ke k

Nilai penduga dari kesalahan standar telah memberikan gambaran ketidakpastian dari penduga tingkat kesalahan. Penduga kesalahan standar merupakan ukuran dari ketidakpastian disekitar kesalahan sebenarnya pada sampel uji atau *data testing* pada pohon regresi saat menghadapi data baru. Breiman (1998:234) menjelaskan bila penduga respon Y apabila X tidak diketahui disimbolkan dengan $\mu = E(Y)$ dan variansi Y atau dikenal dengan rata-rata kuadrat kesalahan (*average square error*) menggunakan nilai μ disimbolkan dengan $R^*(\mu) = E(Y - \mu)^2$.

Diketahui nilai dari $\bar{y} = \frac{1}{N_2} \sum_n y_n$ dan $R(\bar{y}) = \frac{1}{N_2} \sum_n (y_n - \bar{y})^2$ sehingga kesalahan relatif penduga data testing pada pohon ke- k seperti pada persamaan (2.11) yang telah dituliskan kedalam persamaan (2.12) menjadi persamaan seperti berikut:

$$RE^{ts}(T_k) = \frac{R^{ts}(T_k)}{R(\bar{y})} = \frac{\frac{1}{N_2} \sum_{x_n, y_n} (y_n - d_k(x_n))^2}{\frac{1}{N_2} \sum_n (y_n - \bar{y})^2} \quad (2.17)$$

Persamaan (2.17) merupakan persamaan pendugaan rata-rata kuadrat kesalahan dari:

$$RE^*(T_k) = \frac{R^*(T_k)}{R^*\mu} \quad (2.18)$$

$$RE^*(T_k) = \frac{R^*(T_k)}{E(Y - \mu)^2}$$

Penentuan kesalahan standar dari $RE^{ts}(T_k)$ apabila d_k tetap dan pada persamaan (2.18) dapat ditulis dengan:

$$W(V_1, V_2) = \frac{V_1}{V_2} = \frac{\frac{1}{N_2} \sum_n V_{1n}}{\frac{1}{N_2} \sum_n V_{2n}} \quad (2.19)$$

dengan:

$$V_1, V_2 = \text{penduga tetap untuk } \mu_1 = R^*(T_k) \text{ dan } \mu_2 = E(Y - \mu)^2$$

$$W(V_1, V_2) = \text{penduga tetap untuk } W(\mu_1, \mu_2) = \frac{\mu_1}{\mu_2}$$

dengan variansi dari $W(V_1, V_2)$ seperti berikut:

$$\frac{\partial W}{\partial \mu_1} (V_1 - \mu_1) + \frac{\partial W}{\partial \mu_2} (V_2 - \mu_2) = \frac{\mu_1}{\mu_2} \left(\frac{V_1 - \mu_1}{\mu_1} - \frac{V_2 - \mu_2}{\mu_2} \right) \quad (2.20)$$

menghasilkan:

$$\begin{aligned} \text{Var} \left[\frac{\mu_1}{\mu_2} \left(\frac{V_1 - \mu_1}{\mu_1} - \frac{V_2 - \mu_2}{\mu_2} \right) \right] &= \left(\frac{\mu_1}{\mu_2} \right)^2 \text{Var} \left(\frac{V_1 - \mu_1}{\mu_1} - \frac{V_2 - \mu_2}{\mu_2} \right) \\ &= \left(\frac{\mu_1}{\mu_2} \right)^2 \text{Var} \left(\frac{(V_1 - \mu_1)\mu_2 - \mu_1(V_2 - \mu_2)}{\mu_1\mu_2} \right) \\ &= \left(\frac{\mu_1}{\mu_2} \right)^2 \text{Var} \left(\frac{V_1\mu_2 - \mu_1\mu_2 - V_2\mu_1 + \mu_1\mu_2}{\mu_1\mu_2} \right) \\ &= \left(\frac{\mu_1}{\mu_2} \right)^2 \text{Var} \left(\frac{V_1\mu_2 - V_2\mu_1}{\mu_1\mu_2} \right) \\ &= \left(\frac{\mu_1}{\mu_2} \right)^2 \text{Var} \left(\frac{V_1\mu_2}{\mu_1\mu_2} - \frac{V_2\mu_1}{\mu_1\mu_2} \right) \\ &= \left(\frac{\mu_1}{\mu_2} \right)^2 \text{Var} \left(\frac{V_1}{\mu_1} - \frac{V_2}{\mu_2} \right) \\ &= \left(\frac{1}{\mu_1} \right)^2 \text{Var} V_1 + \left(\frac{1}{\mu_2} \right)^2 \text{Var} V_2 \\ &\quad + 2 \left(\frac{1}{\mu_1} \right) \left(-\frac{1}{\mu_2} \right) \text{Covar} (V_1, V_2) \\ &= \frac{\text{Var} V_1}{\mu_1^2} - 2 \frac{\text{Covar} (V_1, V_2)}{\mu_1\mu_2} + \frac{\text{Var} V_2}{\mu_2^2} \\ &= \left(\frac{\mu_1}{\mu_2} \right)^2 \left(\frac{\text{Var} V_1}{\mu_1^2} - \frac{2\text{Cov}(V_1, V_2)}{\mu_1\mu_2} + \frac{\text{Var} V_2}{\mu_2^2} \right) \\ \text{Var} \left[\frac{\mu_1}{\mu_2} \left(\frac{V_1 - \mu_1}{\mu_1} - \frac{V_2 - \mu_2}{\mu_2} \right) \right] &= \frac{1}{N_2} \left(\frac{\mu_1}{\mu_2} \right)^2 \left(\frac{\sigma_1^2}{\mu_1^2} - \frac{2\sigma_{12}}{\mu_1\mu_2} + \frac{\sigma_2^2}{\mu_2^2} \right) \quad (2.21) \end{aligned}$$

dengan:

$$\sigma_1^2 = \text{Var} V_{1n}$$

$$\sigma_2^2 = \text{Var}V_{2n}$$

$$\sigma_{12} = \text{Cov}(U_{1n}, U_{2n}) \forall n \in N_2$$

Misal terdapat data learning sebanyak N_1 yang dipilih secara independen menggunakan distribusi probabilitas tertentu dengan menganggap bahwa *data learning* berguna untuk membentuk sebuah penduga respon $d_{(x)}$. Apabila *data testing* sebanyak N_2 yang dipilih dengan cara independen dari distribusi yang sama dapat dinotasikan dengan $(X_1, Y_1), (X_2, Y_2), \dots, (X_{N_2}, Y_{N_2})$ untuk setiap pasangan (X_n, Y_n) memiliki distribusi yang sama yang mengakibatkan variansi yang dihasilkan sama seperti halnya variansi pasangan yang pertama. Standar deviasi dari R^{ts} adalah sebagai berikut:

$$\text{St. dev } R^{ts} = \frac{1}{\sqrt{N_2}} \{E(Y_1 - d(X_1))^4 - [E(Y_1 - d(X_1))^2]^2\}^{\frac{1}{2}} \quad (2.22)$$

$$\text{Var } R^{ts} = \frac{1}{N_2} \{E(Y_1 - d(X_1))^4 - [E(Y_1 - d(X_1))^2]^2\} \quad (2.23)$$

Kemudian menggunakan estimasi sampel momen sebagai berikut:

- 1) $E(Y_1 - d(X_1))^4 \cong \frac{1}{N_2} \sum_{n=1}^{N_2} (Y_n - d(X_n))^4$
- 2) $E(Y_1 - d(X_1))^2 \cong \frac{1}{N_2} \sum_{n=1}^{N_2} (Y_n - d(X_n))^2 = R^{ts}(d_k)$

dengan menggunakan kedua sampel momen tersebut akan menghasilkan:

$$\begin{aligned} \text{Var } R^{ts} &= \frac{1}{N_2} (E(Y_1 - d(X_1))^4 - [E(Y_1 - d(X_1))^2]^2) \\ &= \frac{1}{N_2} \left(\frac{1}{N_2} \sum_{n=1}^{N_2} (y_n - d(x_n))^4 - (R^{ts}(d_k))^2 \right) \\ &= \frac{1}{N_2^2} \sum_{n=1}^{N_2} (y_n - d(x_n))^4 - \frac{1}{N_2} (R^{ts}(d_k))^2 \end{aligned} \quad (2.24)$$

dengan menggunakan permisalan sebagai berikut:

$$\sigma_1^2 = \text{Var } V_{1n}$$

$$s_1^2 = N_2(\text{Var}(V_1)) \text{ menggunakan } V_1 = R^{ts}(d_k)$$

$$s_1^2 = N_2 \left(\frac{1}{N_2^2} \sum_{n=1}^{N_2} (y_n - \bar{y})^4 - \frac{1}{N_2} (R^{ts}(d_k))^2 \right)$$

$$s_1^2 = \frac{1}{N_2} \sum_{n=1}^{N_2} (y_n - \bar{y})^4 - \frac{1}{N_2} (R^{ts}(d_k))^2 \quad (2.25)$$

Sehingga σ_1^2 dapat diestimasi menggunakan persamaan (2.25). Permisalan selanjutnya dilakukan dengan langkah berikut:

$$\sigma_2^2 = \text{Var } V_{2n}$$

$$s_2^2 = N_2(\text{Var}(V_2)) \text{ menggunakan } V_1 - s^2 = R_{(\bar{y})} = \frac{1}{N_2} \sum_{n=1}^{N_2} (y_n - \bar{y})^2$$

$$\text{St. dev } R_{(\bar{y})} = \frac{1}{N_2} \left\{ E(Y_2 - d(X_2))^4 - \left[E(Y_2 - d(X_2))^2 \right]^2 \right\}^{\frac{1}{2}}$$

$$\text{Var } R_{(\bar{y})} = \frac{1}{N_2} \left(E(Y_2 - d(X_2))^4 - \left[E(Y_2 - d(X_2))^2 \right]^2 \right)$$

Kemudian menggunakan estimasi momen sampel sebagai berikut:

$$1) E(Y_2 - d(X_2))^2 \cong \frac{1}{N_2} \sum_{n=1}^{N_1} (y_n - \bar{y})^4$$

$$2) E(Y_2 - d(X_2))^2 \cong \frac{1}{N_2} \sum_{n=1}^{N_1} (y_n - \bar{y})^2 = R_{(\bar{y})} = s^2$$

dengan menggunakan kedua sampel momen tersebut akan menghasilkan:

$$\begin{aligned} \text{Var } R_{(\bar{y})} &= \frac{1}{N_2} \left(E(Y_2 - d(X_2))^4 - \left[E(Y_2 - d(X_2))^2 \right]^2 \right) \\ &= \frac{1}{N_2} \left(\frac{1}{N_2} \sum_{n=1}^{N_1} (y_n - \bar{y})^4 - (s^2)^2 \right) \\ \text{Var } R_{(\bar{y})} &= \frac{1}{N_2^2} \sum_{n=1}^{N_1} (y_n - \bar{y})^4 - \frac{1}{N_2} s^4 \end{aligned} \quad (2.26)$$

dari persamaan (2.26) diperoleh:

$$\begin{aligned} s_2^2 &= N_2(\text{Var}(V_2)) \\ &= N_2(\text{Var}(R_{(\bar{y})})) \\ &= N_2 \left(\frac{1}{N_2^2} \sum_{n=1}^{N_1} (y_n - \bar{y})^4 - \frac{1}{N_2} s^4 \right) \\ s_2^2 &= \frac{1}{N_2} \sum_{n=1}^{N_1} (y_n - \bar{y})^4 - s^4 \end{aligned} \quad (2.27)$$

Sehingga σ_2^2 dapat diestimasi menggunakan persamaan (2.27). Permisalan selanjutnya dapat dilakukan dengan cara sebagai berikut:

$$\begin{aligned}\sigma_{12} &= Cov(V_{1n}, V_{2n}) \\ s_{12} &= N_2(Cov(V_1, V_2))\end{aligned}$$

$$\begin{aligned}Cov(V_1, V_2) &= \frac{1}{N_2} \left(\frac{1}{N_2} \sum_{n=1}^{N_2} (y_n - d_k(x_n))^2 (y_n - \bar{y})^2 - R^{ts}(d_k) s^2 \right) \\ Cov(V_1, V_2) &= \frac{1}{N_2^2} \sum_{n=1}^{N_2} (y_n - d_k(x_n))^2 (y_n - \bar{y})^2 - \frac{1}{N_2} R^{ts}(d_k) s^2\end{aligned}\quad (2.28)$$

dari persamaan (2.28) akan diperoleh:

$$\begin{aligned}s_{12} &= N_2(Cov(V_1, V_2)) \\ &= N_2 \left(\frac{1}{N_2^2} \sum_{n=1}^{N_2} (y_n - d_k(x_n))^2 (y_n - \bar{y})^2 - \frac{1}{N_2} R^{ts}(d_k) s^2 \right) \\ s_{12} &= \frac{1}{N_2} \sum_{n=1}^{N_2} (y_n - d_k(x_n))^2 (y_n - \bar{y})^2 - R^{ts}(d_k) s^2\end{aligned}\quad (2.29)$$

Sehingga σ_{12} dapat diestimasi menggunakan persamaan (2.29). Persamaan (2.25), (2.27) dan (2.29) sehingga diperoleh rumus dari *standart error* dengan persamaan sebagai berikut:

$$\begin{aligned}SE(R^{ts}(T_k)) &= \sqrt{Var R^{ts}(T_k)} \\ &= \sqrt{\left(\frac{\mu_1}{\mu_2} \right)^2 \left(\frac{Var V_1}{\mu_1^2} - \frac{2Cov(V_1, V_2)}{\mu_1 \mu_2} + \frac{Var V_2}{\mu_2^2} \right)} \\ &= \frac{\mu_1}{\mu_2} \left(\frac{Var V_1}{\mu_1^2} - \frac{2Cov(V_1, V_2)}{\mu_1 \mu_2} + \frac{Var V_2}{\mu_2^2} \right)^{\frac{1}{2}} \\ &= \frac{R^{ts}(T_k)}{R(\bar{y})} \left(\frac{1}{N_2} \left(\frac{s_1^2}{R^{ts}(T_k)^2} \right) - \frac{1}{N_2} \left(\frac{2s_{12}}{R^{ts}(T_k) s^2} \right) + \frac{1}{N_2} \left(\frac{s_2^2}{s^4} \right) \right)^{\frac{1}{2}} \\ SE(R^{ts}(T_k)) &= R^{ts}(T_k) \left[\frac{1}{N_2} \left(\frac{s_1^2}{R^{ts}(T_k)^2} \right) - \left(\frac{2s_{12}}{R^{ts}(T_k) s^2} \right) + \left(\frac{s_2^2}{s^4} \right) \right]^{\frac{1}{2}}\end{aligned}\quad (2.30)$$

2.5 Root Mean Squared Error (RMSE)

Saat mengevaluasi suatu model yang berbeda, diperlukan tingkat ketelitian hasil pemodelan pada penelitian ini menggunakan model kemiskinan Provinsi Jawa Timur. Model terbaik dari CART dan Bagging CART ditentukan dengan

menggunakan *Root Mean Squared Error* (RMSE). Satu keuntungan dari RMSE adalah menghindari penggunaan nilai absolut yang banyak tidak diinginkan dalam perhitungan matematika (Chai, 2014:1249). Nilai RMSE dapat dihitung menggunakan formula berikut:

$$RMSE = \left(\frac{\sum (y_i - \hat{y}_i)}{n} \right)^{\frac{1}{2}} \quad (2.31)$$

$$= \sqrt{\sum_{i=1}^n \left[\frac{(y_i - \hat{y}_i)^2}{n} \right]} \quad (2.32)$$

dengan:

RMSE = *Root Mean Squared Error*

y_i = Nilai prediksi

\hat{y}_i = Nilai aktual

n = Jumlah pengamatan.

Nilai RMSE yang rendah menunjukkan bahwa variansi nilai yang dihasilkan dari suatu model yang diprediksi mendekati variansi nilai observasinya. Semakin kecil nilai RMSE yang dihasilkan, maka semakin dekat nilai yang diprediksi dan diamati dan menunjukkan model terbaik yang dihasilkan.

2.6 Metode Bootstrap

Efron (1979:2) menjelaskan apabila *bootstrapping* merupakan teknik *resampling* statistik yang melibatkan pengambilan sampel acak suatu data set dengan penggantian yang juga sering digunakan sebagai alat mengukur ketidakpastian yang terkait dengan model *machine learning*. Peneliti lainnya, Jason Brownlee (2016) juga mengemukakan apabila bootstrap adalah metode statistik yang kuat untuk memperkirakan kuantitas dari sampel data. Metode ini sangat berguna dikarenakan memungkinkan untuk membuat sampel baru dari suatu populasi tanpa harus mengumpulkan *data learning* tambahan. Bootstrap juga memiliki kegunaan yaitu untuk memperkirakan jumlah statistik dari sampel.

Peningkatan estimasi rata-rata pada metode *bootstrap* dapat dilakukan dengan langkah-langkah sebagai berikut:

- 1) Membuat misalnya 1500 sub-sampel acak dari data set dengan menggunakan penggantian (dapat memilih nilai yang sama untuk beberapa kali).
- 2) Menghitung rata-rata dari setiap sub-sampel yang dihasilkan.
- 3) Menghitung rata-rata dari semua rata-rata yang ada dan menggunakan rata-rata tersebut sebagai rata-rata estimasi pada data.

Proses tersebut dapat digunakan untuk memperkirakan jumlah lain misalnya untuk mengetahui jumlah dari standar deviasi dan jumlah yang digunakan dalam *machine learning*.

2.7 Bootstrap Aggregating (Bagging)

Bootstrap Aggregating atau *Bagging* adalah sebuah prosedur yang memiliki tujuan untuk mengurangi nilai variansi algoritma yang memiliki nilai variansi yang tinggi seperti halnya dari metode CART. *Decision tree* merupakan salah satu metode yang dikombinasikan menggunakan pendekatan *bagging* untuk membentuk estimator yang lebih kuat (Xie, 2019:2). Salah satu cara yang digunakan pada pendekatan *Bootstrap Aggregating* adalah dengan meningkatkan keakuratan hasil yang diperoleh dimana dengan mengambil banyak *data training* dari populasi yang kemudian digunakan untuk membangun prediksi pada setiap model menggunakan *data training*.

Tahap selanjutnya setelah dilakukan pembangunan beberapa model pohon regresi, diperoleh rata-rata prediksi dari masing-masing model yang akan digunakan untuk menentukan prediksi akhir (Kotsiantis, 2006: 54). Breiman (1996:154) dalam penelitiannya memperoleh pengamatan penting apabila kestabilan (responsif terhadap perubahan pada data learning) merupakan prasyarat untuk pendekatan *bagging* agar efektif. Rata-rata pada masing-masing model dapat dihitung menggunakan formula berikut.

$$\hat{f}^1(x), \hat{f}^2(x), \dots, \hat{f}^B(x) \quad (2.33)$$

dengan:

$\hat{f}^1 = \text{bootstrap}$ (pengulangan) ke-1

$\hat{f}^2 = \text{bootstrap}$ (pengulangan) ke-2

$\hat{f}^B = \text{bootstrap}$ (pengulangan) ke- B , dengan B adalah data training terpisah.

Tahap selanjutnya untuk mencari rata-rata untuk menentukan nilai tunggal dengan nilai variansi yang rendah dapat dihitung menggunakan formula berikut.

$$\hat{f}_{ave}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^b(x) \quad (2.34)$$

dengan:

\hat{f}_{ave} = Hasil rata-rata nilai tunggal

\hat{f}^b = Bootstrap (pengulangan) ke- b

Proses *bootstrap* tetap dapat dilakukan meski dengan kondisi B data training yang terpisah. Pengambilan sampel secara berulang set *data learning* tunggal akan menghasilkan B *bootstrap* yang berbeda (Steorts:10). Prediksi dari semua rata-rata yang diperoleh dari melatih metode pada pelatihan bootstrap kedua untuk mendapatkan $f^{*b}(x)$, sehingga formula prediksi rata-rata untuk memperoleh bagging adalah sebagai berikut.

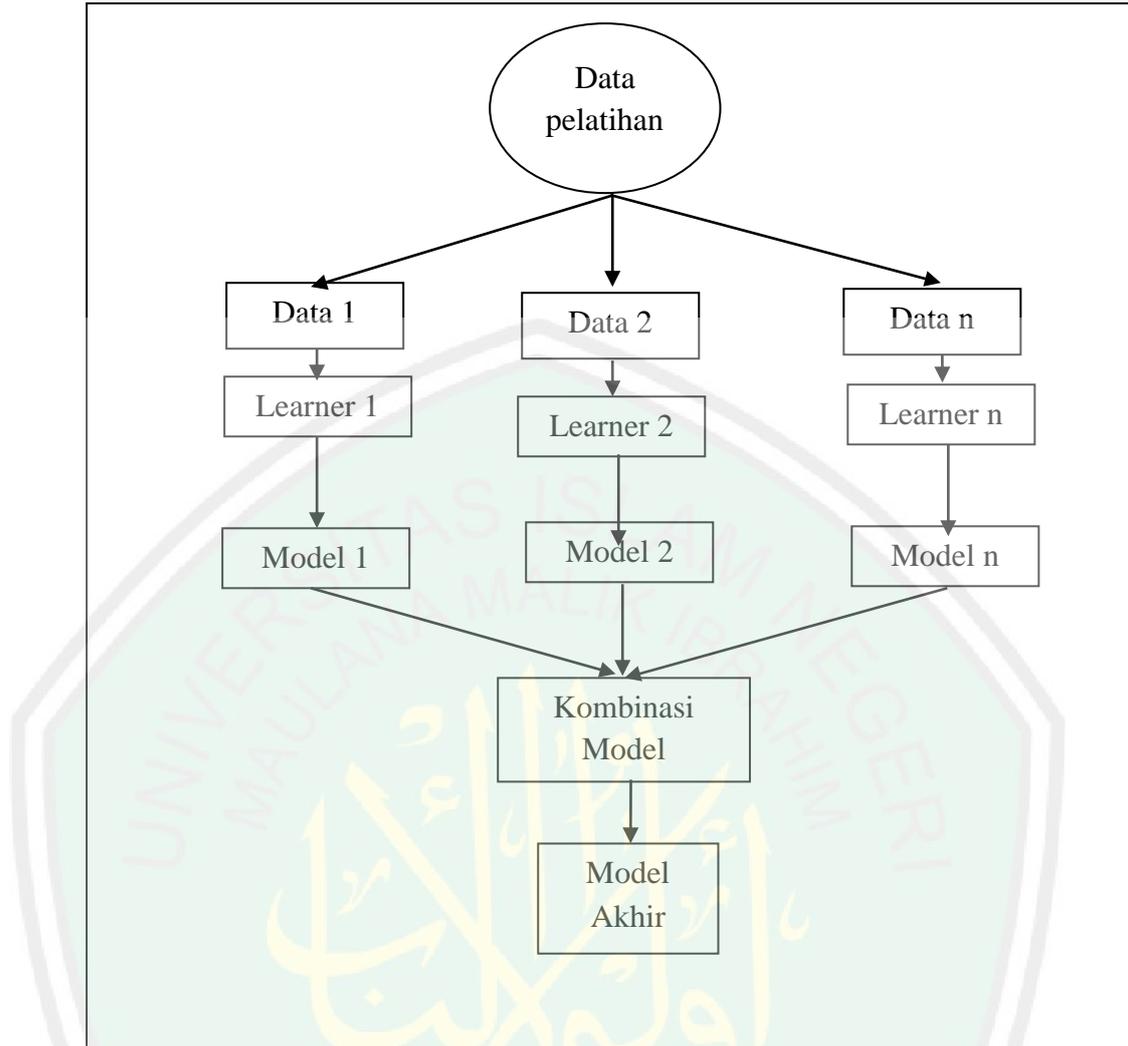
$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^{*b}(x) \quad (2.35)$$

dengan:

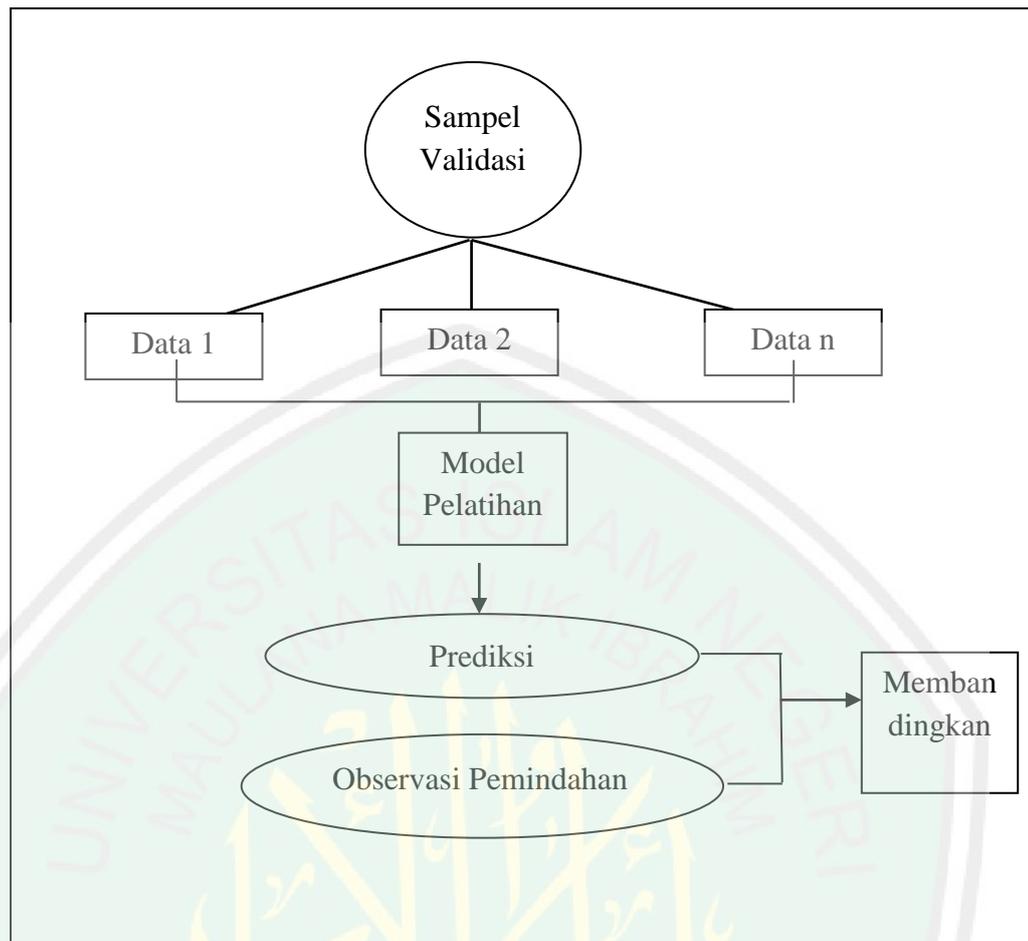
\hat{f}_{bag} = hasil *bagging*

\hat{f}^{*b} = hasil dari *bootstrap* kedua

Rata-rata dari setiap pohon bagging yang diperoleh akan mengambil keuntungan sebesar duapertiga dari pengam atan, dan sepertiga sisanya merupakan model yang tidak cocok dengan model pohon bagging yang diberikan, sepertiga pengamatan tersebut biasa disebut dengan Out Of Bag (OOB). OOB digunakan sebagai data validasi untuk memperkirakan kesalahan prediksi serta menghindari over-fitting yang berlebihan. Kesalahan dari OOB yang telah dihasilkan merupakan perkiraan valid dari kesalahan uji model bagging regresi (Xie, 2019:3).



Gambar 2.4 Prosedur Pelatihan *Bagging*



Gambar 2.5 Prediksi dengan Model *Training*

2.7.1 Langkah-langkah Bootstrap Aggregating (bagging)

Kotsiantis (2006: 55) dalam penelitiannya menjelaskan beberapa langkah-langkah yang harus dilakukan pada pendekatan *bagging*. Langkah-langkah yang dilakukan untuk mendapat keakuratan yang lebih tinggi pada pendekatan *bagging* adalah dilakukan dengan cara sebagai berikut:

- 1) Membangun B pohon regresi dengan menggunakan B *bootstrap* yang berasal dari *data training*.
- 2) Menyimpan model yang telah diperoleh dari proses *bootstrap*.
- 3) Memprediksi nilai kejadian dengan rata-rata dari model yang telah diperoleh dari proses *bootstrap*.

Pohon regresi yang tumbuh serta tanpa dipangkas akan menghasilkan nilai varians yang tinggi serta juga nilai bias yang rendah. Bagging telah memberikan peningkatan yang mengesankan pada keakuratan dengan cara menggabungkan ratusan sampai dengan ribuan pohon menjadi satu prosedur (Steorts: 12).

2.8 Kemiskinan menurut Badan Pusat Statistik (BPS)

Badan Pusat Statistik (BPS) menjelaskan bahwa untuk mengukur kemiskinan konsep yang dilakukan adalah dengan memenuhi kebutuhan dasar. Pemenuhan konsep kebutuhan dasar tersebut akan merubah pola pikir masyarakat yang memandang kemiskinan sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi dasar makanan, dan bukan makanan yang diukur dari sisi pengeluaran. Kemiskinan menjadi topik yang dibahas dan diperdebatkan diberbagai forum nasional maupun di internasional, walaupun kemiskinan telah muncul sejak ratusan tahun yang lalu (Azizah, 2018:168). Masalah kemiskinan dalam penanganannya memerlukan kebijaksanaan yang tepat dalam mengidentifikasi golongan masyarakat yang hidup dibawah garis kemiskinan.

Badan Pusat Statistik (BPS) menerangkan bahwa seseorang masuk kedalam kriteria miskin jika pendapatannya dibawah garis kemiskinan. BPS Indonesia mendasarkan pada besarnya rupiah yang dibelanjakan perkapita/ bulan untuk memenuhi kebutuhan minimum makanan dan non makanan menggunakan patokan 2.100 kalori/ hari kebutuhan non makanan meliputi perumahan, sandang, aneka barang dan jasa. Kemiskinan merupakan suatu fenomena dari hubungan sebab-akibat dimana kemiskinan yang tinggi terjadi ketika rendahnya pendapatan perkapita, pendapatan perkapita yang rendah terjadi karena investasi perkapita yang juga rendah. Tingkat investasi perkapita yang rendah terjadi karena permintaan domestik yang rendah juga dalam hal tersebut menjadikan angka kemiskinan semakin tinggi sehingga membentuk sebuah lingkaran kemiskinan dari sebuah fenomena sebab-akibat.

2.8.1 Kemiskinan di Jawa Timur

Provinsi Jawa Timur termasuk kedalam tiga provinsi dengan tingkat kemiskinan tinggi setelah provinsi DI Yogyakarta dan Jawa tengah pada tahun 2000. Angka kemiskinan Jawa Timur pada tahun 2000 mencapai angka 22,77% yang merupakan tingkat kemiskinan dari 38 kabupaten/kota di Jawa Timur. IPM juga merupakan salah satu keberhasilan suatu penanganan kemiskinan. Kemiskinan yang terjadi pada tahun 2018 di Jawa Timur menurut BPS Provinsi Jawa Timur sudah mencapai angka 10,98% sudah jauh berkurang dibandingkan

dengan tahun 2000. Laju pertumbuhan ekonomi Jawa Timur mempunyai kecenderungan meningkat dari tahun ketahun, namun peningkatan tersebut tidak selalu diiringi oleh penurunan kemiskinan secara signifikan. Jumlah penduduk yang begitu besar dan terus bertambah setiap tahun tidak diimbangi dengan pemerataan penyebaran penduduk (Zamhari, 2015:42).

2.8.2 Variabel yang Berpengaruh Terhadap Kemiskinan di Jawa Timur

Badan Pusat Statistik (BPS) Provinsi Jawa Timur menjelaskan beberapa faktor-faktor pendukung terjadinya kemiskinan di kabupaten dan kota di Jawa Timur, berikut merupakan beberapa faktor kemiskinan:

1. Indeks Pembangunan Manusia (IPM)

IPM merupakan salah satu indikator penting dalam pengukuran keberhasilan dalam upaya pembangunan kualitas hidup manusia atau masyarakat. IPM juga menjadi penentu peringkat pembangunan suatu wilayah dimana bagi Indonesia IPM merupakan data strategis sebagai tolak ukur kinerja dari pemerintah. Dimensi dasar yang dibentuk meliputi umur panjang dan hidup sehat, pengetahuan dan standar hidup yang layak.

Perhitungan IPM mengalami perubahan metodologi, dimana indikator yang berpengaruh pada perubahan metodologi tersebut adalah angka melek huruf pada metode lama digantikan dengan angka harapan lama sekolah dan Produk Domestik Bruto (PDB) perkapita digantikan dengan Produk Nasional Bruto (PNB). Keunggulan dari metode terbaru adalah menggunakan indikator yang lebih tepat dan dapat membedakan dengan baik (diskriminatif) dan menggunakan rata-rata geometrik dalam penyusunan IPM. Berikut merupakan rumus dalam perhitungan IPM:

- Dimensi Kesehatan

$$I_{kesehatan} = \frac{AHH - AHH_{min}}{AHH_{maks} - AHH_{min}}$$

- Dimensi Pendidikan

$$I_{HLS} = \frac{HL\Box - HLS_{min}}{HLS_{maks} - HLS_{min}}$$

$$I_{RLS} = \frac{RLS - RLS_{min}}{RLS_{maks} - RLS_{min}}$$

$$I_{pendidikan} = \frac{I_{HLS} + I_{RLS}}{2}$$

- Dimensi Pengeluaran

$$I_{pengeluaran} = \frac{\ln(\text{pengeluaran}) - \ln(\text{pengeluaran}_{\min})}{\ln(\text{pengeluaran}_{\max}) - \ln(\text{pengeluaran}_{\min})}$$

Sehingga rumus dari IPM adalah sebagai berikut:

$$IPM = \sqrt[3]{I_{kesehatan} \times I_{pendidikan} \times I_{pengeluaran}}$$

2. Produk Domestik Regional Bruto
3. Tenaga Kerja
 - a. Tingkat pengangguran terbuka (TPT) merupakan persentase jumlah pengangguran terhadap jumlah angkatan kerja.
 - b. Tingkat partisipasi Angkatan kerja (TPAK) merupakan persentase jumlah partisipasi angkatan kerja terhadap pengangguran.
4. Rumah tangga pengolah pertanian
5. Pertolongan pertama saat melahirkan
6. Pengeluaran masyarakat
7. Pendidikan yang rendah
Pendidikan yang rendah mengakibatkan seseorang mempunyai ketrampilan tertentu di kehidupannya (Zamzam, 2018:12).
8. Keterbatasan sumberdaya alam
Sumberdaya alam yang terbatas akan mengakibatkan warga terbatas juga dalam pengolahan, sehingga dapat dikatakan masyarakat miskin karena sumberdaya alam nya miskin.
9. Beban keluarga
Memiliki keluarga yang berjumlah banyak apabila tidak diimbangi dengan usaha peningkatan pendapatan akan meimbulkan kemiskinan karena semakin banyak keluarga semakin banyak pula kebutuhan.

2.9 Integrasi Al-Qur'an

Al-Qur'an merupakan sumber pengetahuan dari segala sumber pengetahuan yang ada. 114 surat yang terdapat pada Al-Qur'an tidak ada yang tidak bermanfaat, salah satu contohnya yaitu pada Q.S AL-Kahfi ayat 79:

أَمَّا السَّفِينَةُ فَكَانَتْ لِمَسْكِينٍ يَعْمَلُونَ فِي الْبَحْرِ فَأَرَدْتُ أَنْ أَعِيبَهَا وَكَانَ وَرَاءَهُمْ مَلَكٌ
يَأْخُذُ كُلَّ سَفِينَةٍ غَصْبًا

Q.S Al- Kahfi ayat 79 tersebut memiliki arti: “ Adapun bahtera itu adalah kepunyaan orang-orang miskin yang bekerja dilaut, dan aku bertujuan merusakkan bahtera itu, karena dihadapan mereka ada seseorang raja yang merampas tiap-tiap bahtera”.

Makna yang terkandung dari Q.S Al-Kahfi tersebut adalah apabila orang miskin keadaannya lebih baik daripada orang fakir dikarenakan mereka memiliki perahu atau bahtera yang dapat dijadikan alat untuk mencari nafkah. Ibnu Katsir juga menafsirkan mengenai kemiskinan yang dituangkan dalam Q.S Nisa ayat ke-36 serta juga Allah SWT menegaskan mengenai kemiskinan pada Q.S Al-Balad ayat ke-16. Q.S An-Nisa ayat ke-36 dan Q.S Al-Balad ayat ke-16 memiliki makna apabila yang bernama miskin itu adalah orang yang tidak memiliki apa-apa seperti halnya ucapan Ibnu Katsir “Orang miskin itu adalah orang yang terlantar dan dibuang dijalan”, dimana beliau mengartikan miskin sebagai orang yang tidak memiliki sesuatu yang dapat dibelanjakan (Ibn Katsir, 1998: 2181 dalam Lubis, 2018:71).

Lubis (2018: 71) menjelaskan bahwa menurut tafsir Ibnu Katsir miskin diartikan dengan orang-orang yang membutuhkan yaitu orang-orang yang tidak mendapatkan pihak yang memenuhi kifayahnya, sehingga Allah menyuruh manusia untuk membantu mereka dengan sesuatu yang dapat memenuhi kebutuhannya dan menghilangkan kemudharatannya. Kebutuhan dalam islam menurut Yusuf Qaradhawi bagi semua orang islam yang semestinya tercukupi adalah: jumlah makanan, dan air (HR. Bukhari dan Muslim), pakaian yang menutup aurat (Q.S Al-A’raaf ayat ke-26, Q.S An-Nahl ayat ke-5 dan ayat ke-81), tempat tinggal yang sehat (Q.S An-Nahl ayat ke-80, Q.S An-Nur ayat ke-27), sejumlah harta untuk pernikahan (Q.S An-Nahl ayat ke-72 dan Q.S Ar-Rum ayat ke-21), dan kelebihan harta untuk ibadah haji (Q.S Ali-Imran ayat ke-97) (Lubis, 2018: 76).

Kriteria orang yang tergolong miskin dalam Al-Qur’an menurut Fauzi Arif Lubis (2018) adalah sebagai berikut:

- a. Pihak-pihak yang berhak menerima zakat sebagaimana pada Q.S At-Taubah ayat ke-60.
- b. Orang miskin lebih baik keadaannya dibandingkan dengan orang fakir dikarenakan mereka memiliki perahu atau bahtera yang dapat dijadikan alat untuk mencari nafkah, seperti makna pada Q.S Al-Kahfi ayat ke-79
- c. Miskin walaupun lebih baik dari orang fakir adalah jika orang tersebut tidak mampu untuk memenuhi kebutuhan hidupnya seperti halnya makna dari Q.S Al-Balad ayat ke-16.
- d. Orang miskin adalah orang yang berhak dibantu, sebagaimana ayat yang menyebut orang miskin sebagai pihak yang harus dibantu ekonominya yaitu terdapat dalam 21 ayat yang ada di dalam Al-Qur'an: Q.S Al-Baqarah ayat ke-83, 177, 184, dan 215, Q.S An-Nisa' ayat ke-8 dan 36, Q.S Al-Maidah ayat ke-89 dan 95, Q.S Al-Anfal ayat ke-41, Q.S At-Taubah ayat ke-60, Q.S Al-Isra' ayat ke- 26, Q.S An-Nur ayat ke-22, Q.S Ar-Rum ayat ke-38, Q.S Al-Mujadilah ayat ke- 4, Q.S Al-Hasyr ayat ke-7, Q.S Al-Haqqah ayat ke- 34, Q.S Al-Muddassir ayat ke- 44, Q.S Al-Insan ayat ke-8, Q.S Al-Fajr ayat ke-18, Q.S Al-Balad ayat ke- 16, dan Q.S Al-Ma'un ayat ke-3.

BAB III

METODOLOGI PENELITIAN

3.1 Pendekatan Penelitian

Pendekatan yang digunakan dalam penelitian tentang model kemiskinan di Provinsi Jawa Timur pada tahun 2018 yaitu menggunakan pendekatan studi literatur deskriptif kuantitatif. Penelitian ini merupakan studi penelitian perpustakaan (*Library Research*) yang merupakan penelitian dengan sifat mengumpulkan informasi, mengakumulasikan data dengan menggunakan data kuantitatif atau data numerik.

3.2 Jenis dan Sumber Data

Jenis data yang digunakan pada penelitian ini merupakan data sekunder dimana data yang telah digunakan pada penelitian merupakan data yang sudah ada. Data yang digunakan adalah data tingkat kemiskinan di Provinsi Jawa Timur pada tahun 2018 yang diambil dari Badan Pusat Statistik (BPS) 2019.

3.3 Variabel Penelitian

Variabel pada penelitian ini menggunakan dua variabel, yaitu variabel respon dan juga variabel prediktor. Variabel respon (Y) atau variabel independen merupakan variabel yang dipengaruhi oleh variabel. Variabel prediktor (X) atau variabel dependen merupakan variabel yang mempengaruhi variabel respon. Variabel respon dan variabel prediktor pada penelitian ini yaitu:

Tabel 3.1 Variabel Respon dan Prediktor Kemiskinan Jawa Timur

| Variabel | Deskripsi | Skala |
|----------|---------------------------------------------------------------------------------------------|-------|
| Y | Persentase Kemiskinan (%) | Rasio |
| x_1 | Indeks Pembangunan Manusia (IPM) (%) | Rasio |
| x_2 | Tingkat Partisipasi Angkatan Kerja (TPAK) (%) | Rasio |
| x_3 | Tingkat Pengangguran Terbuka (TPT) (%) | Rasio |
| x_4 | Pengeluaran perkapita (Rupiah) (PP) | Rasio |
| x_5 | Persentase Wanita umur 15-40 pernah kawin dan melahirkan dengan pertolongan dokter (%) (PK) | Rasio |
| x_6 | Rumah tangga usaha pertanian (orang) (SP) | Rasio |
| x_7 | PDRB (%) | Rasio |

3.4 Metode Analisis Data

Langkah-langkah dalam penyelesaian analisis data sebagai berikut:

- 1) Pembentukan statistik deskriptif pada masing-masing variabel yang digunakan.
- 2) Membagi data menjadi dua yaitu *data learning* dan *data testing* dengan proporsi dipilih secara.
- 3) Menganalisis dengan metode *Classification and Regression Trees*
 - a. Melakukan pembentukan pohon pada data learning dimana diperoleh jumlah kuadrat deviasi terbesar digunakan untuk memilah atau menyekat simpul.
 - b. Selanjutnya melakukan penghentian pohon dimana jika tidak lagi terdapat penurunan keheterogenan atau dalam kata lain semua nilai pada simpul y bernilai sama atau homogen.
 - c. Tahap selanjutnya adalah melakukan pemangkasan pada pohon regresi, dimana jika dua anak simpul $\Delta R(s, t) = R(t) - R(t_L) - R(t_R)$ berjumlah 0, maka tahap pemangkasan dimulai dari pohon yang memenuhi hingga berhenti sampai tidak memungkinkan lagi proses pemangkasan, sehingga dihasilkan pemilah terbaik $\Delta R(s^*, t) =$

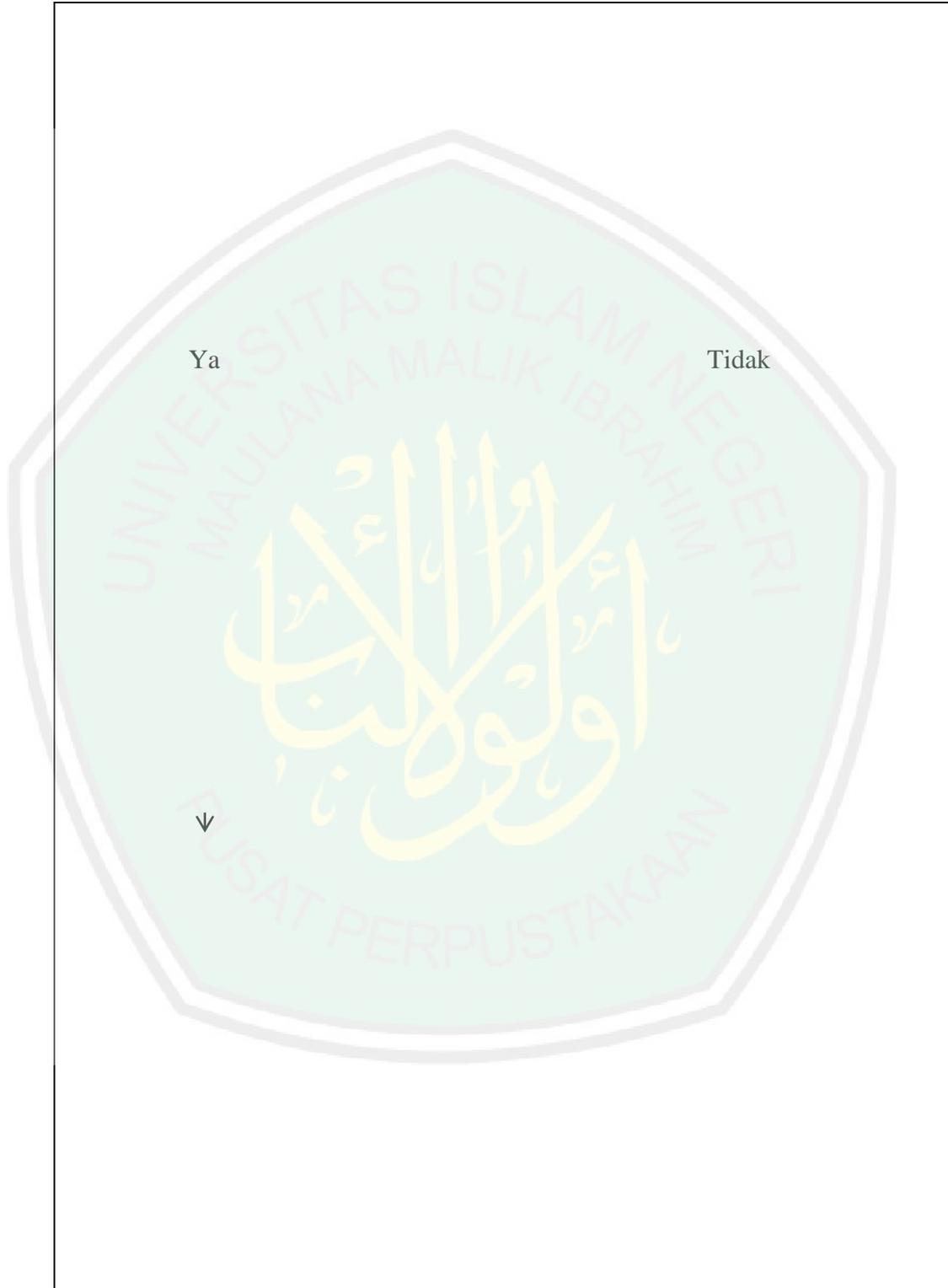
$\max_{s \in S} \Delta R(s, t)$ maka tahap berikutnya adalah pemilihan pohon regresi optimal .

- d. Interpretasi hasil yang diperoleh.
- 4) Melakukan analisis menggunakan pendekatan *Bootstrap Aggregating*.
 - a. Melakukan *Bootstrap Aggregating* dengan menggunakan 50 kali bootstrap (pengulangan) untuk mendapat rata-rata dari setiap model.
 - b. Proses *Aggregat* yaitu penggabungan semua rata-rata model menjadi satu rata-rata untuk memperoleh model terbaik.
 - c. Proses menentukan pohon terbaik dari rata-rata yang telah diperoleh.
 - d. Interpretasi hasil yang diperoleh.
- 5) Menentukan variabel yang paling berpengaruh dalam masalah kemiskinan di Provinsi Jawa Timur tahun 2018.



3.5 Diagram Alur Penelitian

Gambaran dari langkah-langkah analisis data disajikan dalam bentuk diagram alur berikut ini:

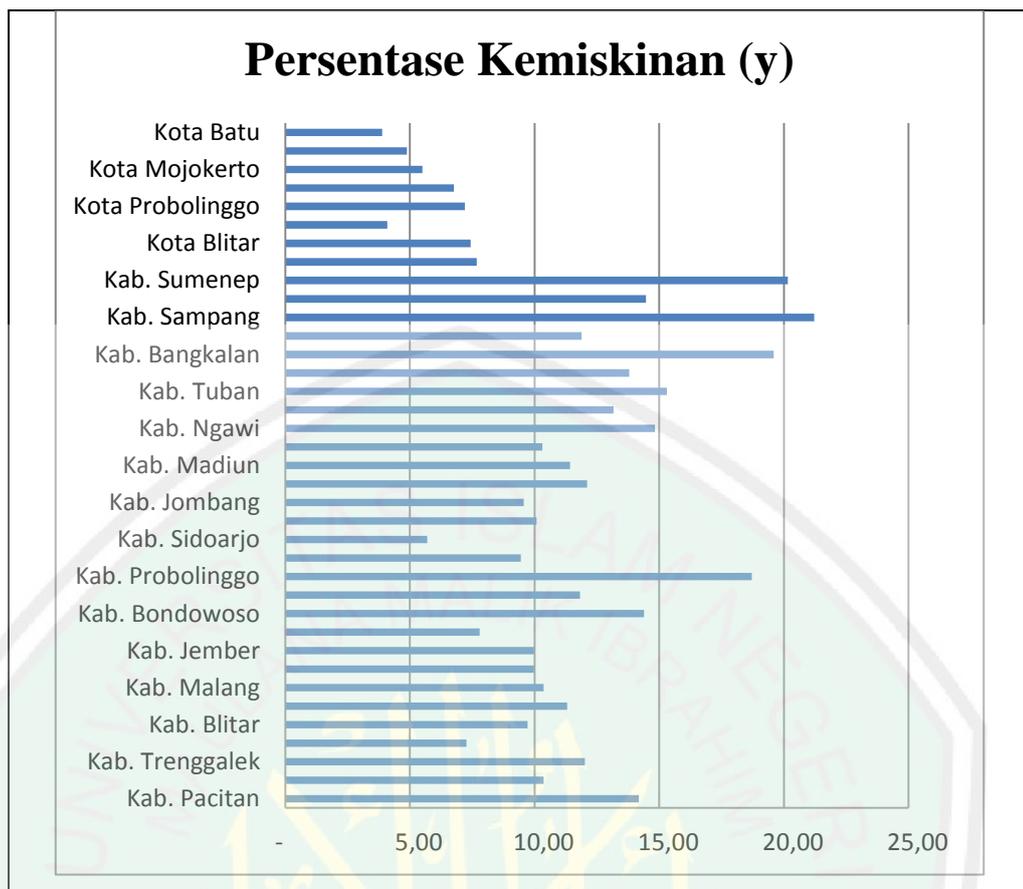


Gambar 3.1 Diagram Alur Pemodelan CART dan *Bagging* CART

BAB IV HASIL DAN PEMBAHASAN

4.1 Deskriptif Data

Kehomogenan data pemilah pada pembentukan pohon regresi dapat dilihat melalui statistik deskriptif. Data penelitian menggunakan data tingkat kemiskinan kabupaten dan kota di Jawa Timur pada tahun 2018 yang akan dimodelkan menggunakan pendekatan metode *Classification and Regression Trees* dan *Bootstrap Aggregating*. Data yang digunakan terdiri dari 38 data tingkat kemiskinan di kabupaten dan kota di Jawa Timur dengan 7 variabel prediktor (x) dan 1 variabel respon (y). Berikut merupakan gambaran umum faktor-faktor penyebab terjadinya kemiskinan di Jawa Timur pada tahun 2018 yang dapat dilihat melalui statistik deskriptif.



Gambar 4.1 Grafik Persentase Kemiskinan di Jawa Timur

Gambar 4.1 merupakan grafik yang menjelaskan tentang persebaran persentase kemiskinan penduduk di Jawa Timur pada tahun 2018. Grafik menunjukkan persentase kemiskinan penduduk Jawa Timur tertinggi berada di daerah Kabupaten Sampang dengan persentase sebesar > 21% lebih tepatnya sebesar 21,21%, dan daerah yang memiliki persentase kemiskinan terendah adalah Kota Batu dengan persentase sebesar < 5% lebih tepatnya 3,89%. Persebaran persentase kemiskinan pada gambar 4.1 tentu saja karena di pengaruhi oleh beberapa faktor sehingga terjadinya kenaikan persentase kemiskinan suatu daerah.

4.1.1 Deskripsi Kemiskinan Provinsi Jawa Timur

Kemiskinan adalah sebuah masalah yang berkaitan dengan yang dapat terjadi dimana saja, baik negara maju maupun berkembang (Prasetyoningrum, 2018: 218). Kemiskinan juga dapat didefinisikan dengan suatu standar hidup yang rendah yang ditandai dengan tingkat kekurangan materi pada sebagian orang

dibandingkan dengan standar hidup orang lain yang umum berlaku dalam masyarakat pada umumnya (Syaifullah, 2017: 110). Deskripsi persentase kemiskinan Provinsi Jawa Timur tahun 2018 berdasarkan kemiskinan yang terjadi pada provinsi Jawa Timur yaitu pada 38 kota/ kabupaten yang ada ditunjukkan pada Tabel 4.1 berikut.

Tabel 4.1 Kemiskinan Provinsi Jawa Timur Tahun 2018

| | | <i>Mean</i> | <i>Median</i> | <i>Maximum</i> | <i>Minimum</i> |
|---|------------|-------------|---------------|----------------|----------------|
| y | Kemiskinan | 10,8861 | 10,34 | 21,21 | 3,89 |

Tabel 4.1 menunjukkan bahwa kemiskinan yang terjadi di Provinsi Jawa Timur pada tahun 2018 memiliki nilai tengah dengan persentase sebesar 10,31. Nilai maksimum yang diperoleh memiliki persentase sebesar 21,21 serta nilai minimum dengan persentase sebesar 3,89. Kemiskinan yang terjadi memiliki nilai rata-rata dengan persentase sebesar 10, 886.

4.1.2 Deskripsi Indeks Pembangunan Manusia (IPM)

Indeks Pembangunan Manusia (IPM) merupakan alat ukur dari pembangunan manusia dimana IPM merupakan suatu indeks komposit untuk mengukur pencapaian kualitas pembangunan manusia (Meriyanti, 2015). Todaro (2006, dalam Meriyanti, 2015) menjelaskan bahwa IPM telah menggambarkan suatu indeks pengembangan manusia yang dilihat dari sisi perluasan, pemerataan, dan keadilan baik dalam bidang kesehatan, pendidikan, maupun kesejahteraan masyarakat. Deskripsi salah satu faktor kemiskinan yang terjadi di Jawa Timur berdasarkan persentase Indeks Pembangunan Manusia (IPM) pada tahun 2018 ditunjukkan pada tabel 4.2 berikut.

Tabel 4.2 IPM Provinsi Jawa Timur Tahun 2018

| | <i>Mean</i> | <i>Median</i> | <i>Maximum</i> | <i>Minimum</i> |
|-------------------------------------------------|-------------|---------------|----------------|----------------|
| Indeks Pembangunan Manusia (IPM) x_1 | 70,9703 | 70,53 | 81,74 | 61,00 |

Tabel 4.2 menunjukkan bahwa kemiskinan dengan kategori indeks pembangunan manusia (IPM) memiliki nilai tengah dengan persentase sebesar 70,53. Sedangkan nilai maksimal dan minimal yang diperoleh dari 38 kota/kabupaten Jawa Timur memiliki nilai persentase sebesar 81,74 dan 61,00. Persentase rata-rata yang diperoleh pada indeks pembangunan manusia (IPM) adalah sebesar 70,9703.

4.1.3 Deskripsi Tingkat Partisipasi Angkatan Kerja (TPAK)

Tingkat partisipasi angkatan kerja (TPAK) merupakan salah satu indikator ketenagakerjaan yang telah memberikan gambaran tentang penduduk yang aktif secara ekonomi dalam kegiatan sehari-hari yang merujuk pada suatu waktu dalam periode survei (Mala, 2017:130). Jumlah penduduk yang semakin banyak dan masuk kedalam golongan bukan angkatan kerja akan mengakibatkan semakin kecilnya jumlah angkatan kerja dan berpengaruh pada semakin kecilnya tingkat partisipasi angkatan kerja (TPAK). Deskripsi salah satu faktor kemiskinan yang terjadi di Jawa Timur berdasarkan persentase tingkat partisipasi angkatan kerja (TPAK) pada tahun 2018 ditunjukkan pada tabel 4.3 berikut.

Tabel 4.3 TPAK Provinsi Jawa Timur Tahun 2018

| | <i>Mean</i> | <i>Median</i> | <i>Maximum</i> | <i>Minimum</i> |
|----------------------------------------------------------|-------------|---------------|----------------|----------------|
| Tingkat Partisipasi Angkatan Kerja (TPAK) x_2 | 69,6926 | 69,44 | 79,41 | 64,41 |

Tabel 4.2 menunjukkan bahwa kemiskinan Provinsi Jawa Timur yang dipengaruhi oleh tingkat partisipasi angkatan kerja (TPAK) memiliki persentase nilai tengah sebesar 69,44. Persentase nilai maksimal dan nilai minimal dari tingkat partisipasi angkatan kerja diperoleh sebesar 79,41 dan 64,41. Sedangkan persentase rata-rata yang diperoleh adalah sebesar 69,692.

4.1.4 Deskripsi Tingkat Pengangguran Terbuka (TPT)

Tingkat pengangguran terbuka merupakan persentase jumlah pengangguran terhadap jumlah angkatan kerja (Astuti, 2017: 152). Tingkat pengangguran terbuka juga merupakan kunci dari kinerja perekonomian yang ada (Priastiwi, 2019: 160). Kegunaan dari tingkat pengangguran terbuka sendiri adalah untuk mengindikasikan besarnya persentase angkatan kerja yang termasuk dalam pengangguran. Deskripsi salah satu faktor kemiskinan yang terjadi di Jawa Timur berdasarkan persentase tingkat pengangguran terbuka (TPT) pada tahun 2018 ditunjukkan pada tabel 4.4 berikut.

Tabel 4.4 TPT Jawa Timur Tahun 2018

| | <i>Mean</i> | <i>Median</i> | <i>Maximum</i> | <i>Minimum</i> |
|------------------------------------|-------------|---------------|----------------|----------------|
| Tingkat Pengangguran Terbuka (TPT) | 3,7832 | 3,840 | 6,79 | 1,43 |
| x_3 | | | | |

Tabel 4.4 menunjukkan bahwa tingkat pengangguran terbuka (TPT) yang terjadi di Provinsi Jawa Timur memiliki persentase nilai tengah sebesar 3,840. Persentase nilai maksimal serta minimal yang diperoleh dari pengangguran terbuka adalah sebesar 6,79 dan 1,43. Sedangkan rata-rata pengangguran terbuka memiliki persentase sebesar 3,7832.

4.1.5 Deskripsi Pengeluaran Perkapita (Riil yang Disesuaikan) Provinsi Jawa Timur

Muda (2019: 48) menjelaskan apabila pengeluaran perkapita merupakan alat pengukur standar hidup dari manusia. Pengeluaran perkapita merupakan salah satu faktor penyebab kemiskinan apabila masyarakat selalu bergaya hidup konsumtif. Pengetahuan serta peluang yang ada untuk merealisasikan pengetahuan dalam berbagai kegiatan produktif sehingga mendapatkan hasil output yang baik berupa barang ataupun jasa sebagai pendapatan merupakan salah satu pengaruh pada pengeluaran perkapita. Pengeluaran perkapita telah memberikan gambaran mengenai tingkat daya beli PPP (Purchasing Power Parity) masyarakat, serta sebagai salah satu komponen yang digunakan dalam melihat

status pembangunan manusia pada suatu wilayah (Yunita, 2012 dalam Muda ,2019:48). Deskripsi salah satu faktor kemiskinan yang terjadi di Jawa Timur berdasarkan pengeluaran perkapita dengan satuan rupiah pada tahun 2018 ditunjukkan pada tabel 4.5 berikut.

Tabel 4.5 Pengeluaran Perkapita (Rupiah) Provinsi Jawa Timur tahun 2018

| | <i>Mean</i> | <i>Median</i> | <i>Maximum</i> | <i>Minimum</i> |
|----------------|-------------|---------------|----------------|----------------|
| Pengeluaran | | | | |
| Perkapita (PP) | 11.202,5 | 10.926 | 17.157 | 8.393 |
| x_4 | | | | |

Berdasarkan tabel 4.5 menunjukkan bahwa pengeluaran perkapita (PP) dengan satuan rupiah memiliki nilai tengah sebesar Rp 10.926. Nilai maksimal dan minimal yang diperoleh adalah sebesar Rp 17.157 dan Rp 8.393. sedangkan rata-rata pengeluaran perkapita Provinsi Jawa Timur adalah sebesar Rp 11.202,5.

4.1.6 Deskripsi Wanita Umur 15-40 Pernah Kawin dan Melahirkan Hidup dengan Pertolongan Dokter Provinsi Jawa Timur (PK)

Angka harapan hidup merupakan sebuah alat untuk mengevaluasi kinerja pemerintah dalam meningkatkan kesejahteraan penduduk pada umumnya dan meningkatkan derajat esehatan pada umumnya (muda: 2019: 47). Kelahiran merupakan lahirnya janin berusia 5 bulan keatas dengan menggunakan penolong kelahiran. Deskripsi salah satu faktor kemiskinan yang terjadi di Jawa Timur berdasarkan persentase wanita umur 15-40 yang pernah kawin dan melahirkan bayi hidup pada tahun 2018 ditunjukkan pada tabel 4.7 berikut.

Tabel 4. 6 Persentase Melahirkan Hidup Provinsi Jawa Timur Tahun 2018

| | <i>Mean</i> | <i>Median</i> | <i>Maximum</i> | <i>Minimum</i> |
|------------|-------------|---------------|----------------|----------------|
| Persentase | | | | |
| Kelahiran | 37,72 | 35,57 | 63,37 | 14,42 |
| x_5 | | | | |

Berdasarkan tabel 4.6 menunjukkan bahwa persentase nilai tengah yang diperoleh dari persentase melahirkan hidup dengan penolong dokter adalah sebesar 35,57. Persentase nilai maksimal yang diperoleh sebesar 63,37 yang

berada diatas 50% yang dapat meningkatkan angka harapan hidup Provinsi Jawa Timur serta persentase minimum yang didapat sebesar 14,42. Persentase melahirkan hidup memperoleh persentase rata-rata sebesar 37,72.

4.1.7 Deskripsi Jumlah Rumah Tangga Usaha Pertanian (Jenis Usaha Pertanian Utama yang Diusahakan, 2018) Provinsi Jawa Timur (SP)

Sektor pertanian memiliki keterkaitan yang kuat dengan eksistensi kemiskinan. Wilayah Asia Tenggara merupakan negara dimana sektor pertanian menjadi kontribusi terhadap *Product Domestic Bruto* (PDB) sebanyak 10% dan telah menyediakan lapangan pekerjaan lebih dari sepertiga jumlah penduduknya (Fan and Zuang, 2009 dalam Hermawan, 2012:135). Deskripsi salah satu faktor kemiskinan yang terjadi di Jawa Timur berdasarkan sektor pertanian dengan satuan per-individu pada tahun 2018 ditunjukkan pada tabel 4.7 berikut.

Tabel 4.7 Sektor Pertanian (Individu) Provinsi Jawa Timur Tahun 2018

| | <i>Mean</i> | <i>Median</i> | <i>Maximum</i> | <i>Minimum</i> |
|------------------|-------------|---------------|----------------|----------------|
| Sektor Pertanian | 13120 | 13331 | 41210 | 242 |
| x_6 | | | | |

Tabel 4.7 menunjukkan bahwa nilai tengah dari masyarakat yang masih berada di lingkup sektor pertanian sebanyak 13120 individu. Sektor pertanian Provinsi Jawa Timur menginjak nilai maksimal sebesar 41210 serta nilai minimal sebesar 242 individu penduduk yang berada pada lingkup sektor pertanian. Sedangkan nilai rata-rata yang diperoleh adalah sebanyak 13120 individu.

4.1.8 Deskripsi PDRB (Atas Dasar Harga Konstan 2010) Provinsi Jawa Timur

Pertumbuhan ekonomi yang tinggi dan berkelanjutan merupakan kondisi utama atau keharusan bagi kelangsungan pembangunan ekonomi dan peningkatan kesejahteraan (Dama, 2016:556). Apabila pertumbuhan ekonomi tanpa dibarengi penambahan lapangan pekerjaan akan menyebabkan ketimpangan dalam pembagian dan penambahan pendapatan yang akan menciptakan suatu kondisi dimana pertumbuhan ekonomi dengan peningkatan kemiskinan. PDRB atas dasar

harga konstan digunakan untuk mengetahui pertumbuhan ekonomi dari tahun ke tahun, sedangkan menurut BPS PDRB atas dasar harga berlaku digunakan untuk menunjukkan besarnya struktur perekonomian dan peranan sektor ekonomi (Susanti, 2013:5). Deskripsi salah satu faktor kemiskinan yang terjadi di Jawa Timur berdasarkan persentase PDRB pada tahun 2018 ditunjukkan pada tabel 4.8 berikut.

Tabel 4.8 PDRB Provinsi Jawa Timur Tahun 2018

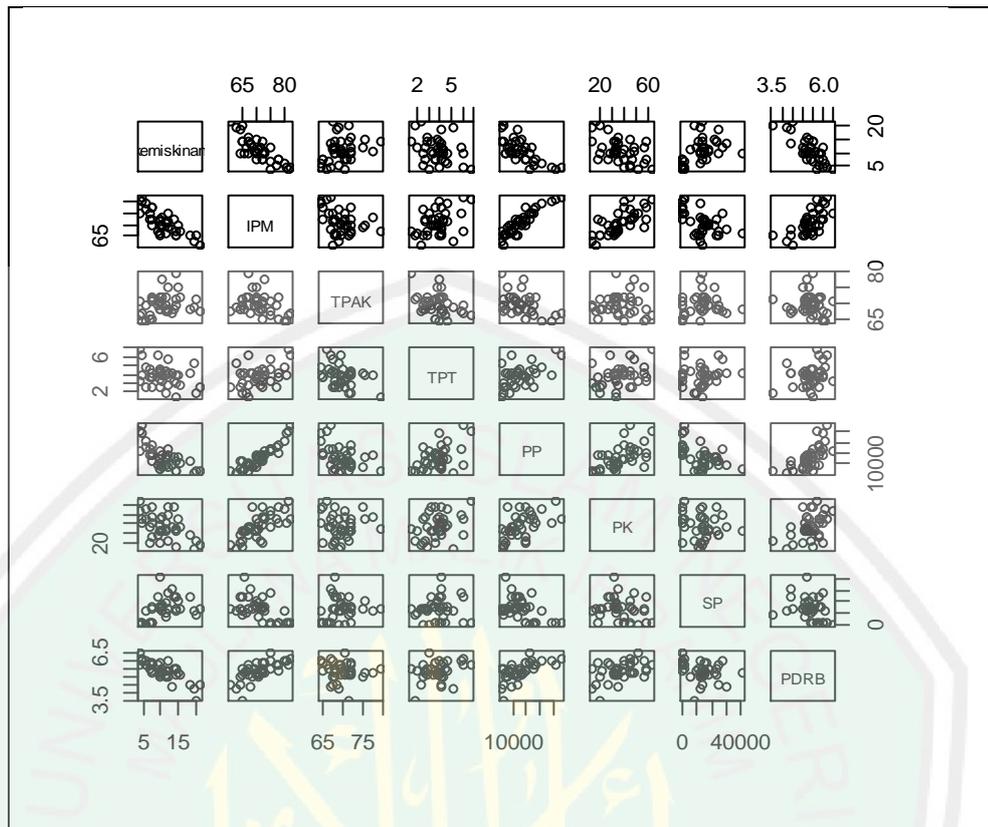
| | <i>Mean</i> | <i>Median</i> | <i>Maximum</i> | <i>Minimum</i> |
|-------|-------------|---------------|----------------|----------------|
| PDRB | 5,3610 | 5 | 6,5 | 3,58 |
| x_7 | | | | |

4.2 Pemodelan Kemiskinan Provinsi Jawa Timur Menggunakan Metode CART dan *Bagging* CART

Proses awal untuk melakukan analisis pemodelan kemiskinan Provinsi Jawa Timur menggunakan metode CART dan *Bagging* CART adalah dengan membagi data menjadi dua yaitu *data learning* dan *data testing*. Kegunaan *data learning* sendiri adalah pondasi untuk membangun model pohon regresi yang diinginkan, dan kegunaan dari *data testing* adalah untuk memvalidasi model dengan mengetahui seberapa besar kemampuan model dalam memprediksi data baru (Pratiwi, 2014). Penelitian ini menggunakan perbandingan 90:10 pada pembagian data, dimana 90% untuk *data learning* yang digunakan yaitu sebanyak 34 data dan 10% untuk *data testing* yaitu sebanyak 4 data. Penentuan pohon terbaik dapat dilihat berdasarkan Nilai *RMSE* terkecil yang diperoleh pada masing-masing model.

4.2.1 Model Classification And Regression Trees (CART)

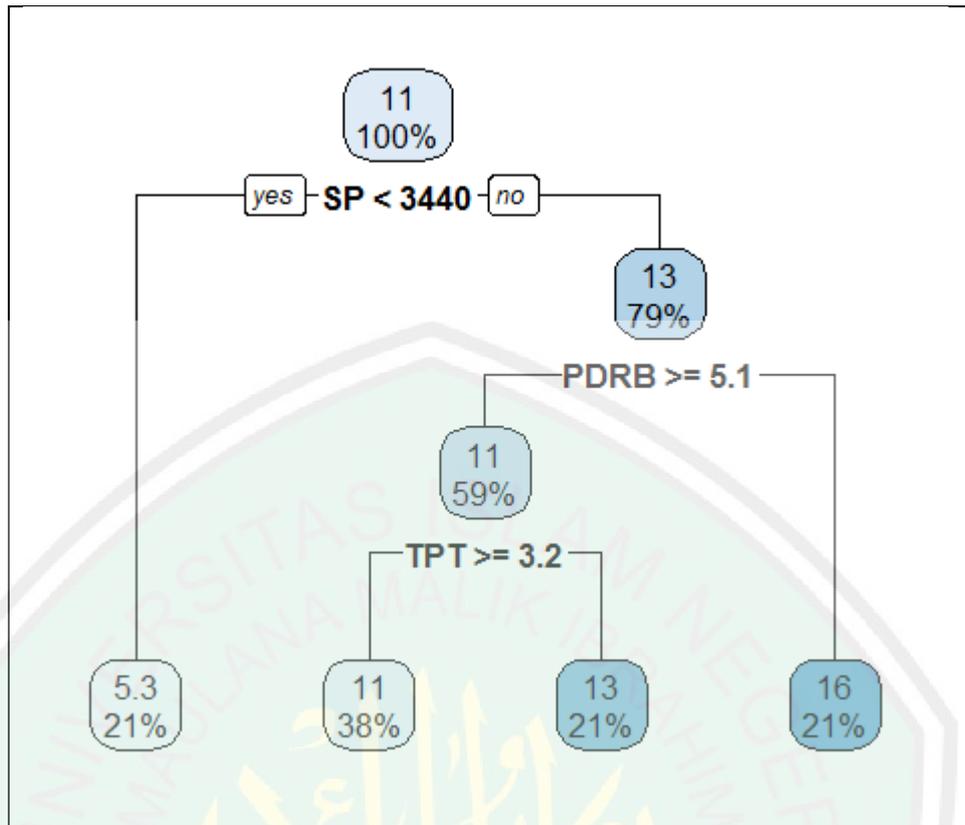
Penelitian ini diketahui mengenai metode non parametrik yaitu *Classification And Regression Trees* (CART) dan *Bagging* CART dikarenakan masing-masing variabel memiliki pola yang berbeda-beda, pola tidak beraturan dan tidak jelas sehingga sulit untuk mengetahui hubungan yang terjadi antar variabel. Gambar 4. Merupakan gambaran dari pola yang tidak beraturan pada data yang digunakan pada penelitian kali ini.



Gambar 4.2 Plot *Matrix* Kemiskinan Provinsi Jawa Timur

Proses awal dari metode *Classification And Regression Trees* adalah pembangunan pohon regresi, pembangunan pohon regresi yang dilakukan dalam penelitian ini menurut acuan dari aturan pemilahan dan aturan pertumbuhan pohon dengan memulai data kemiskinan Jawa Timur dari 38 sample dengan satu variabel respon dan tujuh variabel prediktor. Pertumbuhan pohon dimulai pada data kemiskinan terkumpul dalam suatu himpunan yang disebut dengan simpul akar dan diberi nama simpul 1. Masing-masing kemungkinan dari pemilahan akan menghasilkan dua kelompok data yaitu simpul anak kanan dan simpul anak kiri yang dinamai dengan simpul 2 dan simpul 3.

Pemilihan yang dilakukan diharapkan dapat memaksimalkan ukuran kehomogenan pada masing-masing simpul anak relatif terhadap simpul induk serta memaksimalkan ukuran pemisahan antara simpul anak kiri dan simpul anak kanan yang telah terbentuk. Berikut merupakan pohon regresi yang dihasilkan pada tahap pembangunan pohon regresi.

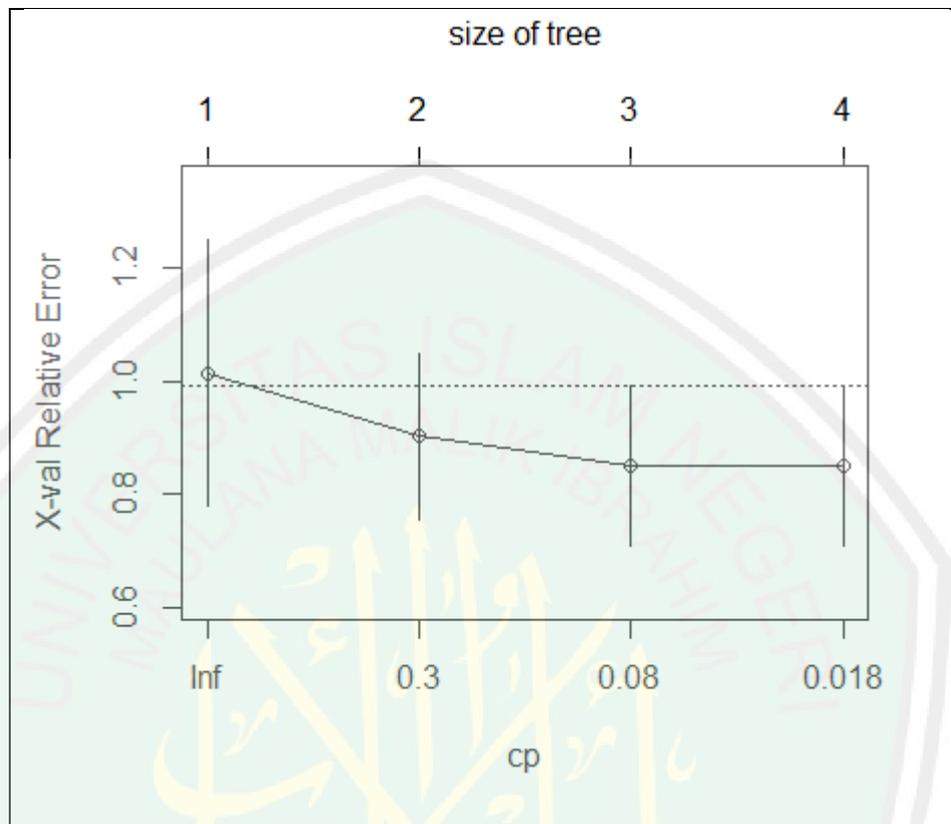


Gambar 4.3 Output Pohon Regresi Maksimal

Berdasarkan Gambar 4.3 proses pembangunan pohon regresi dan hasil plot terlihat bahwa terdapat 6 simpul akhir pada pohon maksimal yaitu:

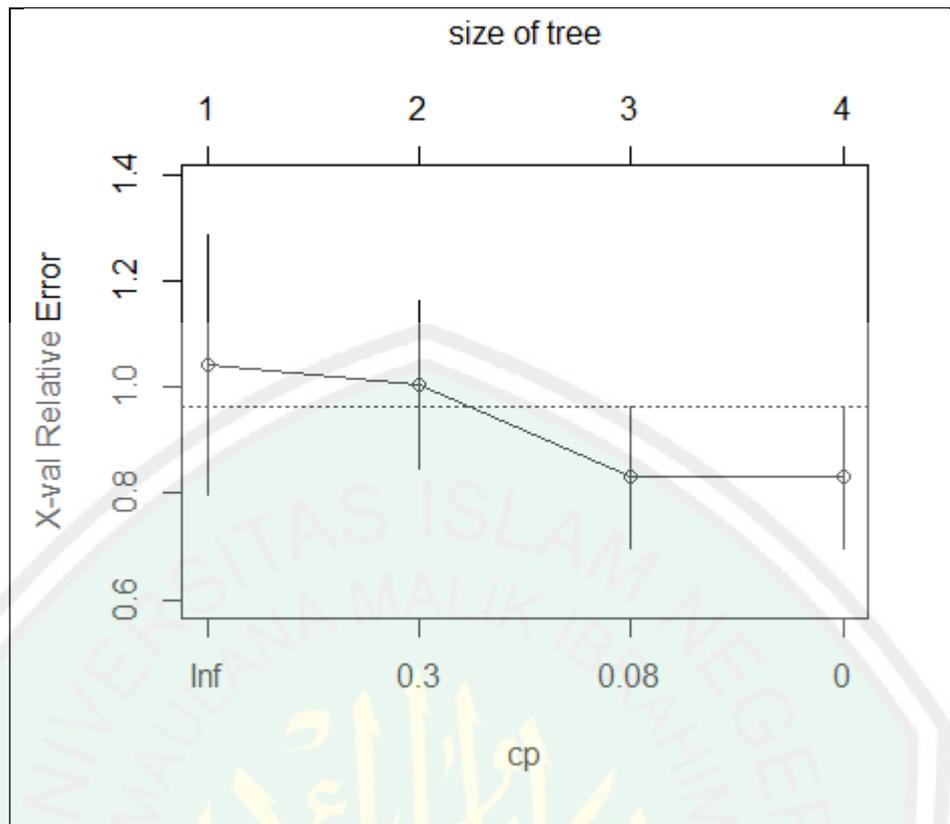
1. Kelompok amatan yang tergolong variabel $SP < 3440$ memiliki nilai rata-rata 5,261429 sebesar 21% atau sebanyak 7 data.
2. Kelompok amatan yang tergolong variabel $SP \geq 3440$ memiliki nilai rata-rata 12,625190 sebesar 79% atau sebanyak 27 data.
3. Kelompok amatan yang tergolong variabel $PDRB \geq 5,055$ memiliki nilai rata-rata 11,302500 sebesar 21% atau sebanyak 20 data.
4. Kelompok amatan yang tergolong variabel $TPT \geq 3,205$ memiliki nilai rata-rata 10,544620 sebesar 21% atau sebanyak 13 data.
5. Kelompok amatan yang tergolong variabel $TPT < 3,205$ memiliki nilai rata-rata 12,710000 sebesar 38% atau sebanyak 7 data.
6. Kelompok amatan yang tergolong variabel $PDRB < 5,055$ memiliki nilai rata-rata 16,404290 sebesar 59% atau sebanyak 7 data.

Proses selanjutnya adalah pemangkasan pohon regresi menggunakan *crossvalidation-10* (validasi silang-10) yang digambarkan melalui plot berikut:



Gambar 4.4 Plot Parameter Kompleksitas Pemangkasan (cp)

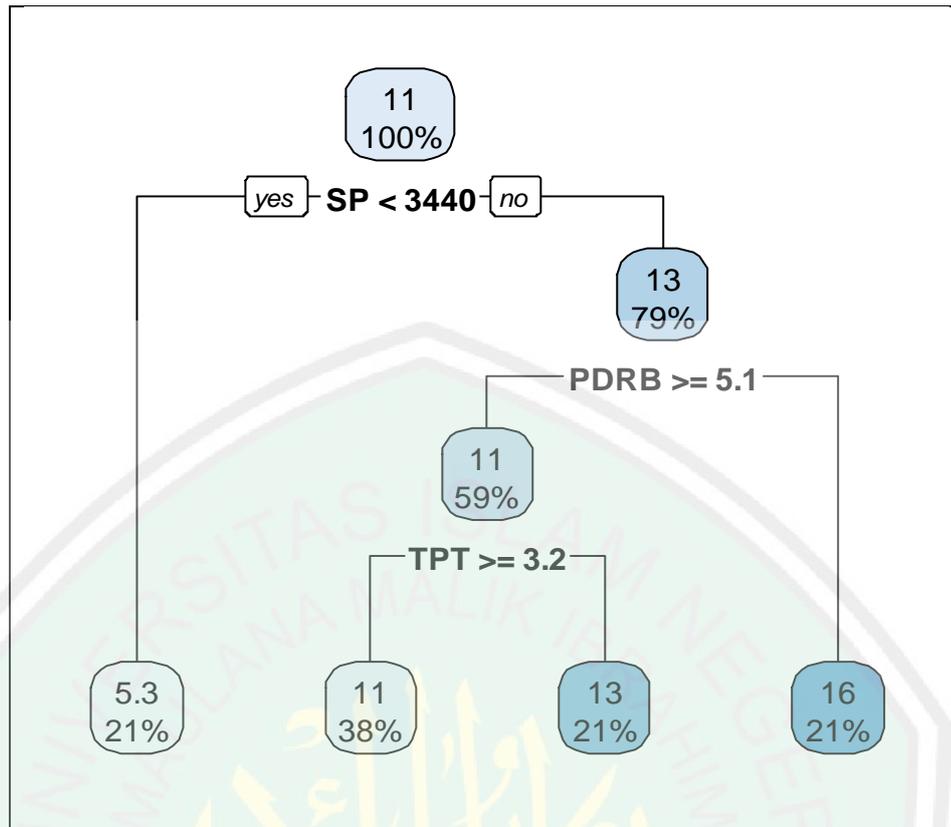
Gambar 4.4 menjelaskan bahwa semakin banyak pohon yang terbentuk semakin kecil pula kesalahan yang dihasilkan. Cp menggambarkan kesalahan validasi silang relatif (sumbu y) dan berbagai nilai cp (sumbu x), dimana nilai cp yang lebih kecil mengarah pada pohon yang lebih besar (sumbu x atas). Hal ini menunjukkan bahwa biaya kompleksitas (*crossvalidation*) yang dibutuhkan untuk pemangkasan akan semakin kecil, sehingga kita dapat memangkas pohon secara signifikan dan masih mencapai kesalahan minimal yang diharapkan. Menggunakan ukuran pohon 4 memberikan hasil validasi silang yang optimal.



Gambar 4.5 Plot Parameter Kompleksitas Pemangkasan (cp) Pohon Dewasa

Gambar 4.5 menjelaskan bahwa dengan *subtree optimal* sebesar 3 *splits* dan 4 *node* terminal, kesalahan *crossvalidation-10* yang dihasilkan adalah sebesar 0,851 dengan kesalahan relatif 0,314. Peningkatan kinerja model dapat dilakukan dengan *hyperparameter*, dalam penelitian ini mencari rentang *minsplit* 2-20 dan *maxdepth* 4 – 15 dikarenakan model asli penelitian ini menemui pohon optimal 4 dan hasilnya adalah sebanyak 228 kombinasi yang berbeda, yang membutuhkan 228 model berbeda. Setelah melakukan filter untuk 5 nilai kesalahan minimal teratas, model optimal membuat sedikit peningkatan daripada model yang sebelumnya yaitu *xerror* 0,513 vs 0,851.

Proses sebelumnya telah mendapatkan hasil apabila menggunakan ukuran pohon 4 telah menghasilkan validasi silang yang optimal, setelah dilakukan validasi silang Gambar 4.5 merupakan model pohon optimal yang dihasilkan.



Gambar 4.6 Output Pohon Regresi Optimal

4.2.2 Model *Bootstrap Aggregating*

Proses *Bootstrap Aggregating* merupakan proses yang akan dilakukan apabila pada metode *Classification And Regression Trees* telah menemukan pohon optimal dengan model terbaiknya. *Bootstrap Aggregating* dilakukan dengan tujuan untuk mengetahui perbandingan hasil terbaik dengan nilai keakuratan tertinggi yang dilihat dari nilai kesalahan pada model terbaik yang dihasilkan pada metode *Classification And Regression Trees* dan *Bootstrap Aggregating*. Secara umum pendekatan *Bootstrap Aggregating* melakukan 25 sample *bootstrap* akan tetapi pada penelitian ini menggunakan bootstrapping sebanyak 100 kali replikasi untuk pencarian model terbaik dengan nilai kesalahan yang sangat rendah.

Pendekatan *Bootstrap Aggregating* menggunakan $\frac{2}{3}$ dari data sebagai data *training* dan $\frac{1}{3}$ data merupakan data yang keluar dari sampel *bootstrap* atau disebut dengan *Out Of Bag* (OOB). OOB akan digunakan untuk memperkirakan keakuratan dari model serta juga untuk menciptakan proses validasi silang yang

alami. Pendekatan *Bootstrap Aggregating* memiliki banyak manfaat antara lain lebih mudah untuk melakukan *cross validation-10* (validasi silang), akan tetapi pada penelitian ini kesalahan juga diperoleh melalui proses OOB. Proses *cross validation-10* akan memberikan pemahaman yang lebih kuat mengenai kesalahan pada pengujian dan juga dapat menilai pentingnya variabel yang digunakan pada *Bootstrap Aggregating*. Proses *cross validation-10* pada penelitian ini telah dilakukan dan memperoleh nilai *Mean Absolute error* (MAE) sebesar 2,180356 dan R^2 sebesar 0,7012645.

Lampiran 10 menunjukkan hasil apabila dilakukan penambahan pohon pada plot *RMSE* pada *Bootstrap Aggregating* regresi yang akan diilustrasikan pada lampiran 13. Kemiskinan Provinsi Jawa Timur dengan menggunakan pendekatan *Bootstrap Aggregating* memperoleh model dengan mendapatkan manfaat paling banyak dari agregasi dalam *Bootstrap Aggregating* seperti yang diilustrasikan pada lampiran 13.

Setelah dibandingkan antara uji yang telah ditetapkan oleh sampel dengan *RMSE* yang telah melalui proses *cross validation-10* hasilnya cukup dekat, sehingga pada proses *Bootstrap Aggregating* ini telah berhasil mengurangi kesalahan dari model terbaik yang didapat.

4.2.3 Perbandingan Keakuratan Model

Berdasarkan hasil yang telah diperoleh pada *Classification And Regression Trees* dan *Bootstrap Aggregating* maka perbandingan tingkat keakuratan yang diukur dengan menggunakan *RMSE* yang telah sesuai dengan perbedaan rata-rata antara nilai yang telah dihasilkan pada pengamatan dan nilai yang diperoleh dari prediksi model (Kassambara, 2017:176). Berikut merupakan nilai *RMSE* yang telah dihasilkan dari kedua metode.

Tabel 4.9 Perbandingan Keakuratan metode *Classification And Regression Trees* dan *Bootstrap Aggregating*

| | <i>RMSE</i> |
|--------------------------------------------|-------------|
| <i>Classification And Regression Trees</i> | 3,498211 |
| <i>Bootstrap Aggregating</i> | 2,881789 |

Berdasarkan tabel 4.1 diketahui bahwa nilai kesalahan yang dihasilkan melalui metode *Classification And Regression Trees* adalah sebesar 3,498211 dan nilai kesalahan yang diperoleh dari metode *Bootstrap Aggregating* adalah sebesar 2,881789. Nilai *RMSE* yang dihasilkan pada masing-masing model menunjukkan bahwa model dari *Bootstrap Aggregating* memiliki nilai kesalahan paling rendah. *RMSE* menunjukkan bahwa model *Bootstrap Aggregating* memiliki keakuratan lebih tinggi dibandingkan metode *Classification And Regression Trees*. Perbandingan model *RMSE* antara *Classification And Regression Trees* dan *Bootstrap Aggregating* disajikan pada lampiran 12 dan lampiran 13.

4.3 Variabel yang mempengaruhi kemiskinan Provinsi Jawa Timur menurut metode *Classification And Regression Trees* dan *Bootstrap Aggregating*.

Variabel yang berpengaruh atau *variable importance* adalah tingkat kepentingan dari suatu variabel yang digunakan pada penelitian untuk mengetahui interaksi antara variabel-variabel dalam membentuk sebuah model. Penyebab kemiskinan yang terjadi di Provinsi Jawa Timur pada tahun 2018 antara lain disebabkan oleh beberapa variabel yang dapat dilihat berdasarkan *variable importance* yang telah dihasilkan saat penelitian. Berdasarkan model yang telah dihasilkan, terdapat urutan kepentingan dari masing-masing variabel yang mempengaruhi terjadinya kemiskinan di Jawa Timur. Berikut merupakan urutan dari *variable importance* yang telah dihasilkan:

Tabel 4.10 Variable Importance Classification And Regression Trees

| Overall | |
|---------|-------|
| PDRB | 100 |
| IPM | 97,24 |
| PP | 82,62 |
| SP | 81 |
| PK | 61,67 |
| TPT | 0 |
| TPAK | 0 |

Model kemiskinan Provinsi Jawa Timur tahun 2018 menggunakan *Classification And Regression Trees* telah digambarkan dengan gambar 4.5. Lampiran 7 menjelaskan bahwa pada metode *Classification And Regression Trees* untuk model kemiskinan di Provinsi Jawa Timur tahun 2018 variabel yang mempengaruhi kemiskinan tersebut dalam skala interval adalah PDRB (Produk Domestik Regional Bruto). *Variable Importance* pada urutan kedua yaitu adalah variabel IPM (Indeks Pembangunan Manusia). Selanjutnya pada urutan ketiga yaitu variabel PP (Pengeluaran Perkapita), pada urutan keempat yaitu variabel SP (Rumah tangga usaha pertanian), urutan kelima yaitu variabel PK (Persentase wanita umur 15-40 tahun pernah kawin dan melahirkan dengan pertolongan dokter), urutan keenam yaitu variabel TPT (Tingkat Pengangguran Terbuka) dan urutan selanjutnya yaitu variabel TPAK (Tingkat Partisipasi Angkatan Kerja).

Selanjutnya, telah diketahui juga model yang dihasilkan dari metode *Bootstrap Aggregating*, sehingga *variable importance* dari metode *Bootstrap Aggregating* yang mempengaruhi terjadinya kemiskinan di Provinsi Jawa Timur adalah sebagai berikut:

Tabel 4.11 Variable Importance Bootstrap Aggregating

| Overall | |
|---------|--------|
| IPM | 100 |
| PP | 83,445 |
| TPT | 76,990 |
| TPAK | 75,589 |
| PK | 53,765 |
| PDRB | 4,453 |
| SK | 0 |

Kemiskinan Provinsi Jawa Timur dengan menggunakan metode *Bootstrap Aggregating* memperoleh model dengan mendapatkan manfaat paling banyak dari agregasi dalam *Bootstrap Aggregating* seperti yang diilustrasikan pada lampiran 13. Variabel yang paling penting pada pemodelan kemiskinan dengan skala interval 100 dengan metode *Bootstrap Aggregating* sendiri pada urutan pertama

yaitu variabel IPM (Indeks Pembangunan Manusia). *Variable importance* pada urutan selanjutnya yaitu variabel PP (Pengeluaran perkapita), urutan ketiga yaitu variabel TPT (Tingkat Paengangguran Terbuka), kemudian pada urutan keempat yaitu variabel TPAK (Tingkat Partisipasi Angkatan Kerja), urutan kemima yaitu variabel PK (Persentase Kelahiran dengan bantuan dokter), selanjutnya pada urutan keenam yaitu variabel PDRB (Produk Domestik Regional Bruto) dan diurutan terakhir terdapat variabel SP (Sektor Pertanian per-individu).



BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan pada analisis pembahasan, dapat ditarik kesimpulan sebagai berikut:

1. Keakuratan dari model yang dihasilkan pada data kemiskinan Provinsi Jawa Timur tahun 2018 menggunakan metode *Classification And Regression Trees* memperoleh nilai *RMSE* sebesar 3,498211. Selanjutnya, keakuratan model juga dihitung menggunakan metode *Bootstrap Aggregating* dengan perolehan nilai *RMSE* sebesar 2,881789. Sehingga, dapat disimpulkan bahwa dalam penelitian ini khususnya pada data kemiskinan Provinsi Jawa Timur tahun 2018 setelah dilakukan perbandingan keakuratan menggunakan nilai *RMSE* diketahui apabila pada *Bootstrap Aggregating* memperoleh nilai *RMSE* lebih kecil dari pada metode *Classification And Regression Trees*. Jadi, pada penelitian ini metode *Bootstrap Aggregating* adalah metode yang tepat digunakan untuk mengetahui model terbaik dari data kemiskinan Provinsi Jawa Timur karena memiliki tingkat keakuratan yang lebih tinggi dibandingkan dengan metode *Classification And Regression Trees*.
2. Variabel yang paling mempengaruhi untuk terjadinya kemiskinan di Provinsi Jawa Timur tahun 2018 menggunakan metode *Classification And Regression Trees* pada urutan pertama yaitu variabel PDRB yang diukur dengan skala interval memperoleh nilai sebesar 100,00. Selanjutnya, variabel yang berpengaruh terhadap kemiskinan di Provinsi Jawa Timur tahun 2018 menurut *Bootstrap Aggregating* pada urutan pertama yang juga diukur dengan skala interval memperoleh nilai sebesar 100,00 pada variabel IPM.

5.2 Saran

Penelitian ini menggunakan metode *Classification And Regression Trees* Dan *Bootstrap Aggregating* khususnya pohon regresi dalam mencari keakuratan model terbaik serta variabel yang sangat mempengaruhi pada data penelitian.

Diharapkan penelitian selanjutnya mengembangkan penelitian ini dengan menggunakan *software* analisis lainnya. Penelitian selanjutnya diharapkan menambah kelompok variabel-variabel lainnya yang dalam penelitian ini belum digunakan dalam pembangunan pohon regresi dan bagging pohon regresi.



DAFTAR PUSTAKA

- Astuti, W.I., dkk. 2017. *Analisis Faktor Yang Berpengaruh Terhadap Tingkat Pengangguran Terbuka Di Provinsi Jawa Timur menggunakan Regresi Panel*. 6(1): 150-156.
- Azizah, dkk. 2018. *Pengaruh Pendidikan, Pendapatan Perkapita dan Jumlah Penduduk Terhadap Kemiskinan di provinsi Jawa Timur*. 2(1): 167-180.
- Badan Pusat Statistik. Ekonomi Dan Perdagangan. Produk Domestik Regional Bruto (Kabupaten/ Kota) . *Laju Pertumbuhan PDRB Atas Dasar Harga Konstan 2010 Provinsi Jawa Timur Menurut Kabupaten/ kota (persen), 2014-2018* (Online). (<https://jatim.bps.go.id/statictable/2019/10/09/1712/laju-pertumbuhan-pdrb-atas-dasar-harga-konstan-2010-provinsi-jawa-timur-menurut-kabupaten-kota-persen-2014-2018.html>). Diakses 1 Februari 2020.
- Badan Pusat Statistik. Sosial Dan Kebudayaan. Kesehatan. Distribusi Persentase Wanita Berumur 15-49 Tahun Yang Pernah Kawin Dan Melahirkan Hidup Dalam Dua Tahun Terakhir Menurut Kabupaten/ Kota Dan penoong Persalinan Di Jawa Timur, 2017 Dan 2018 (Online). (<https://jatim.bps.go.id/statictable/2019/10/09/1672/distribusi-persentase-wanita-berumur-15-49-tahun-yang-pernah-kawin-dan-melahirkan-hidup-dalam-dua-tahun-terakhir-menurut-kabupaten-kota-dan-penolong-persalinan-di-jawa-timur-2017-dan-2018-.html>). Diakses 1 Februari 2020.
- Badan Pusat Statistik. Sosial Dan Kependudukan. *Indeks pembangunan Manusia. Indeks Pembangunan Manusia Di Provinsi Jawa Timur Menurut Kabupaten/ Kota, 2013-2018* (Online). (<https://jatim.bps.go.id/statictable/2019/10/11/1861/indeks-pembangunan-manusia-di-provinsi-jawa-timur-menurutkabupaten-kota-2013-2018-.html>). Diakses 1 Februari 2020.
- Badan Pusat Statistik. Sosial Dan Kependudukan. Konsumsi Dan Pengeluaran. *Pengeluaran Per Kapita Per Tahun Yang Disesuaikan Menurut Kabupaten/ Kota Di Jawa Timur* (Online). (<https://jatim.bps.go.id/statictable/2019/10/09/1643/tingkat-pengangguran-terbuka-tpt-dan-tingkat-partisipasi-angkatan-kerja-tpak-menurut-kabupaten-kota-2016-2018.html>). Diakses 1 Februari 2020.

- Badan Pusat Statistik. Sosial Dan Kependudukan. Tenaga Kerja. *Tingkat Pengangguran Terbuka (TPT) Dan Tingkat Partisipasi Angkatan Kerja (TPAK) Menurut Kabupaten/ Kota, 2016-2018 (Online)*. (<https://jatim.bps.go.id/statistictable/2019/10/09/1643/tingkat-pengangguran-terbuka-tpt-dan-tingkat-partisipasi-angkatan-kerja-tpak-menurut-kabupaten-kota-2016-2018.html>). Diakses 1 Februari 2020.
- Badan Pusat Statistik. Pertanian Dan Pertambangan. Tanaman Pangan. *Jumlah Rumah Tangga Usaha Pertanian Dengan Sumber Penghasilan Utama Dari Sektor Pertanian Menurut Kabupaten/ Kota Dan Jenis Usaha Pertanian Utama Yang Diusahakan, 2018 (Online)*. (<https://jatim.bps.go.id/statistictable/2019/10/17/2062/jumlah-rumah-tangga-usaha-pertanian-dengan-sumber-penghasilan-utama-dari-sektor-pertanian-menurut-kabupaten-kota-dan-jenis-usaha-pertanian-utama-yang-diusahakan-2018.html>). Diakses 1 Februari 2020.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. California. U.S.A: Wodsworth, Inc.
- Brownlee. Jason., 2016. Bagging And Random Forest Ensemble Algorithms For Machine Learning. Web: <https://machinelearningmastery.com/bagging-and-random-forest-ensemble-algorithms-for-machine-learning/>
- Cahya, Bayu Tri. 2015. Kemiskinan Ditinjau dari Perspektif Al-Qur'an dan Hadist. *Jurnal Penelitian*. 9(1): 41-66.
- Chai. T., & Draxier.R. R., 2014. *Root Mean Squared Error (RMSE) or Mean Absolute Error (MAE)?- Arguments Against Avoiding RMSE In The Literature*. 7: 1247- 1250.
- Dama, Himawan Yudhistira., dkk. 2016. *Pengaruh Produk Domestik Regional Bruto (PDRB) Terhadap Tingkat Kemiskinan Di Kota Manado (Tahun 2011-2014)*. 16(3): 549-561.
- Dewi, K.W, H.S., dkk. 2012. *Analisis Faktor-Faktor Yang Mempengaruhi Indeks Mutu Benang Menggunakan Pohon Regresi*. 5(2): 75-86.
- Diao, Liqun., & Weng, Chengguo., 2019. *Regression Tree Credibility Model*. 23(2):169-196.
- Efron. B. 1979. *Bootstrap Methods: Another Look At Jackknife*. 7(1): 1-26.
- Eynaud,Y., dkk., 2012. *Toward A Simplification Of Models using Regression Trees*. 10:1-14.
- Hermawan, Iwan. 2012. *Analisis Eksistensi Pertanian Terhadap Pengurangan Kemiskinan Di Pedesaan & Perkotaan*. 28(2): 135-144.

- Kassambara, Alboukadel. 2017. *Machine Learning Essentials*. STHDA.
- Komalasari, W.B., 2007. *Metode Pohon Regresi Untuk Eksploratori data Dengan Peubah Yang Banyak Dan Kompleks*. 16(1): 967-980
- Kotsiantis, S.B., dkk. 2006. *Bagged Averaging Of Regression Models*. University Of Patras, Greece. 204: 53-60.
- Lewis, R. J., 2000 . *An introduction to classification and regression tree (CART) analysis*. In *Annual meeting of the society for academic emergency medicine in San Francisco, California*.
- Lubis, Fauzi Arif. 2018. *Miskin Menurut Pandangan Al-Qur'an*. Tansiq. 1(1): 68-82.
- Mala, V.S.N., dkk. 2017. *Analisis Tingkat Partisipasi Angkatan Kerja Berdasarkan Kegiatan Ekonomi Masyarakat Desa Tegalsari Kecamatan Tegalsari Kabupaten Banyuwangi Tahun 2015*. 11(1): 130-139.
- Meriyanti, Ni Komang. 2015. *Pengaruh Program Indeks Pembangunan Manusia (IPM) Terhadap Pengentasan Kemiskinan Di Kecamatan Buleleng Tahun 2011-2014*. 5(1).
- Muda, Riyan., dkk. 2019. *Pengaruh Angka Harapan Hidup, Tingkat Pendidikan Dan Pengeluaran Perkapita Terhadap Pertumbuhan Ekonomi Di Sulawesi Utara Pada Tahun 2003-2017*. 19(1): 44-55.
- Munir, Rinaldi. 2010. *Variansi Dan Kovariansi*. Web:<https://informatika.stei.itb.ac.id/~rinaldi.munir/Probstat/2010-2011/probstat10-11.htm#SlideKuliah>. Dari Bahan Kuliah//2092 Probabilitas Dan Statistik.
- Prasad, M., Anantha., 2006. *Newer Classification And Regression Trees Techniques: Bagging And Random Forest For Ecological Prediction*. 9: 181-199.
- Prasetyoningrum, Ari Kristin. 2018. *Analisis Pengaruh Indeks Pembangunan Manusia, pertumbuhan Ekonomi dan Pengangguran Terhadap Kemiskinan Di Indonesia*. 6(2): 217-240.
- Priastiwi, Dian., & Handayani, Herniwati Retno. 2019. *Analisis Pengaruh Jumlah Penduduk, Pendidikan, Upah Minimum, dan PDRB Terhadap Tingkat Pengangguran Terbuka Di Provinsi Jawa Tengah*. 1(1): 159-269.
- Ratnaningrum, dkk., 2016. *Analisis Klasifikasi Nasabah Kredit Menggunakan Bootstrap Aggregating And Regression Trees (Bagging Cart)*. 5(1): 81-90.

- Ratnasari,V., & Qurratu'ain,A,Q., 2016. *Indikator Tingkat Kemiskinan di Jawa Timur menggunakan Regresi Panel*. 3(2): 2337-3520.
- Suharjo, Budi., 2008. *Metode Pohon Regresi Dan Prosedur Regresi Bertatar Untuk Segmentasi Data*. 7(1): 39-54.
- Susanti, Sussy. 2013. *Pengaruh Produk Domestik Regional Bruto, Pengangguran dan Indeks Pembangunan Manusia Terhadap Kemiskinan Di Jawa Barat Dengan Menggunakan Analisis Data Panel*. 9(1): 1-18
- Steorts, C., Rebecca., Tree Based Methods: Bagging, Bosting, And Regression Trees. Duke University. Web: http://www2.stat.duke.edu/~rcs46/lectures_2017/08-trees/08-tree-advanced.pdf. STA 325, Chapter 8 ISL.
- Syaifullah, Ahmad., & Malik, Nazarudin. 2017. *Pengaruh Indeks Pembangunan Manusia (IPM) Dan Pruduk Domestik Bruto (PDB) Terhadap Tingkat Kemiskinan Di ASEAN-4 (Studi 4 Negara ASEAN)*. 1(1): 107-119.
- Xie, Jiemin., dkk., 2019. *Ensemble Of Bagged Regression Tree For Concrete Dam Deformation Predicting*. 376: 1-7.
- Yohannes, Y., & Webb, P., 1999. *Classification And Regression Trees, CART™ A User Manual For Identifying Indicator Of Valnulerability To Famine And Chronic Food Insecurity*. Washington, D,C: International Food Policy Research Institute.
- Zamzam, dkk., 2018. *Faktor Yang Mempengaruhi Tingkat Kemiskinan Di Desa Madello Kecamatan Balusu Kabupaten Barru*. (Doctoral Disertation. Universitas Negeri Makasar).

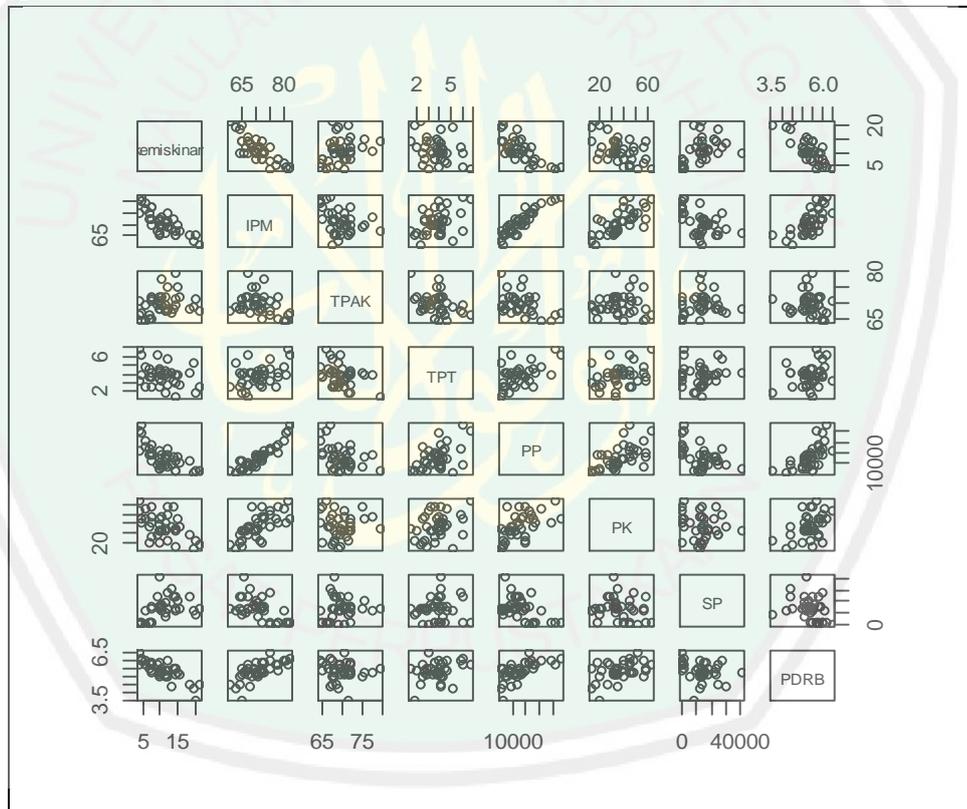
LAMPIRAN

Lampiran 1. Data Prosentase Kemiskinan Provinsi Jawa Timur Tahun 2018

| | y | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Kab. Pacitan | 14,19 | 67,33 | 79,41 | 1,43 | 8527 | 34,00 | 13704 | 5,51 |
| Kab. Ponorogo | 10,36 | 69,91 | 72,07 | 3,87 | 9426 | 32,66 | 15270 | 5,31 |
| Kab. Trenggalek | 12,02 | 68,71 | 75,19 | 4,17 | 9400 | 56,60 | 13144 | 5,03 |
| Kab. Tulungagung | 7,27 | 71,99 | 70,62 | 2,61 | 10455 | 46,95 | 14046 | 5,21 |
| Kab. Blitar | 9,72 | 69,93 | 70,61 | 3,37 | 10327 | 33,39 | 15396 | 5,10 |
| Kab. Kediri | 11,31 | 71,07 | 67,70 | 4,25 | 10853 | 40,49 | 5807 | 5,08 |
| Kab. Malang | 10,37 | 69,40 | 69,70 | 3,24 | 9844 | 33,35 | 17471 | 5,56 |
| Kab. Lumajang | 9,98 | 64,83 | 68,10 | 2,55 | 8931 | 20,79 | 12750 | 5,02 |
| Kab. Jember | 9,98 | 65,96 | 67,90 | 4,09 | 9090 | 32,60 | 41210 | 5,23 |
| Kab. Banyuwangi | 7,80 | 70,06 | 71,12 | 3,67 | 11828 | 26,97 | 16128 | 5,84 |
| Kab. Bondowoso | 14,39 | 65,27 | 71,45 | 3,90 | 10429 | 32,28 | 23250 | 5,09 |
| Kab. Situbondo | 11,82 | 66,42 | 71,87 | 1,92 | 9692 | 33,24 | 9668 | 5,43 |
| Kab. Probolinggo | 18,71 | 64,85 | 68,41 | 4,15 | 10700 | 27,85 | 24199 | 4,47 |
| Kab. Pasuruan | 9,45 | 67,41 | 69,59 | 6,11 | 9933 | 28,50 | 24758 | 5,79 |
| Kab. Sidoarjo | 5,69 | 79,50 | 64,53 | 4,73 | 14168 | 56,13 | 11909 | 6,05 |
| Kab. Mojokerto | 10,08 | 72,64 | 71,92 | 4,27 | 12454 | 45,59 | 15128 | 5,85 |
| Kab. Jombang | 9,56 | 71,86 | 69,86 | 4,64 | 10999 | 39,20 | 14893 | 5,43 |
| Kab. Nganjuk | 12,11 | 71,23 | 67,91 | 2,64 | 11768 | 22,76 | 14359 | 5,39 |
| Kab. Madiun | 11,42 | 71,01 | 69,52 | 3,81 | 11351 | 33,75 | 18183 | 5,10 |
| Kab. Magetan | 10,31 | 72,91 | 77,60 | 3,92 | 11539 | 47,81 | 11610 | 5,25 |
| Kab. Ngawi | 14,83 | 69,91 | 75,41 | 3,83 | 11187 | 45,49 | 21341 | 5,26 |
| Kab. Bojonegoro | 13,16 | 67,85 | 67,13 | 4,19 | 9926 | 35,81 | 30276 | 4,41 |
| Kab. Tuban | 15,31 | 67,43 | 71,78 | 2,83 | 10048 | 35,72 | 13493 | 5,16 |
| Kab. Lamongan | 13,8 | 71,97 | 68,02 | 3,17 | 11108 | 58,46 | 23828 | 5,50 |
| Kab. Gresik | 11,89 | 75,28 | 67,29 | 5,82 | 12845 | 48,88 | 26731 | 5,97 |
| Kab. Bangkalan | 19,59 | 62,87 | 68,86 | 5,25 | 8393 | 14,42 | 12410 | 4,26 |
| Kab. Sampang | 21,21 | 61,00 | 67,31 | 2,41 | 8569 | 17,11 | 13169 | 4,51 |
| Kab. Pamekasan | 14,47 | 65,41 | 69,35 | 2,92 | 8536 | 14,82 | 11009 | 5,46 |

| | | | | | | | | |
|------------------|-------|-------|-------|------|-------|-------|------|------|
| Kab. Sumenep | 20,16 | 65,25 | 71,53 | 1,79 | 8722 | 21,16 | 8312 | 3,58 |
| Kota Kediri | 7,68 | 77,58 | 65,09 | 3,63 | 11976 | 58,37 | 424 | 5,42 |
| Kota Blitar | 7,44 | 77,58 | 72,21 | 4,06 | 13391 | 40,33 | 616 | 5,83 |
| Kota Malang | 4,10 | 80,89 | 65,94 | 6,79 | 16158 | 63,37 | 1073 | 5,72 |
| Kota Probolinggo | 7,20 | 72,53 | 64,89 | 3,64 | 11796 | 19,85 | 462 | 5,94 |
| Kota Pasuruan | 6,77 | 74,78 | 66,33 | 4,55 | 12931 | 35,85 | 723 | 5,54 |
| Kota Mojokerto | 5,50 | 77,14 | 69,19 | 2,45 | 13155 | 44,91 | 295 | 5,80 |
| Kota Madiun | 4,49 | 80,33 | 64,41 | 3,85 | 15616 | 40,24 | 398 | 5,96 |
| Kota Batu | 3,89 | 75,04 | 70,52 | 3,12 | 12466 | 52,56 | 242 | 6,50 |
| Kota Surabaya | 4,88 | 81,74 | 66,98 | 6,12 | 17157 | 45,94 | 865 | 6,20 |

Lampiran 2. Matrix Plot



Lampiran 3. Output Data Learning

| | kemiskinan | IPM | TPAK | TPT | PP | PK | SP | PDRB |
|---|------------|-------|-------|------|-------|-------|-------|------|
| 1 | 14.19 | 67.33 | 79.41 | 1.43 | 8527 | 34.00 | 13704 | 5.51 |
| 2 | 10.36 | 69.91 | 72.07 | 3.87 | 9426 | 32.66 | 15270 | 5.31 |
| 3 | 12.02 | 68.71 | 75.19 | 4.17 | 9400 | 56.60 | 13144 | 5.03 |
| 4 | 7.27 | 71.99 | 70.62 | 2.61 | 10455 | 46.95 | 14046 | 5.21 |
| 5 | 9.72 | 69.93 | 70.61 | 3.37 | 10327 | 33.39 | 15396 | 5.10 |

| | | | | | | | | |
|----|-------|-------|-------|------|-------|-------|-------|------|
| 6 | 11.31 | 71.07 | 67.70 | 4.25 | 10853 | 40.49 | 5807 | 5.08 |
| 7 | 10.37 | 69.40 | 69.70 | 3.24 | 9844 | 33.35 | 17471 | 5.56 |
| 8 | 9.98 | 64.83 | 68.10 | 2.55 | 8931 | 20.79 | 12750 | 5.02 |
| 9 | 9.98 | 65.96 | 67.90 | 4.09 | 9090 | 32.60 | 41210 | 5.23 |
| 10 | 7.80 | 70.06 | 71.12 | 3.67 | 11828 | 26.97 | 16128 | 5.84 |
| 12 | 11.82 | 66.42 | 71.87 | 1.92 | 9692 | 33.24 | 9668 | 5.43 |
| 13 | 18.71 | 64.85 | 68.41 | 4.15 | 10700 | 27.85 | 24199 | 4.47 |
| 14 | 9.45 | 67.41 | 69.59 | 6.11 | 9933 | 28.50 | 24758 | 5.79 |
| 16 | 10.08 | 72.64 | 71.92 | 4.27 | 12454 | 45.59 | 15128 | 5.85 |
| 17 | 9.56 | 71.86 | 69.86 | 4.64 | 10999 | 39.20 | 14893 | 5.43 |
| 18 | 12.11 | 71.23 | 67.91 | 2.64 | 11768 | 22.76 | 14359 | 5.39 |
| 19 | 11.42 | 71.01 | 69.52 | 3.81 | 11351 | 33.75 | 18183 | 5.10 |
| 20 | 10.31 | 72.91 | 77.60 | 3.92 | 11539 | 47.81 | 11610 | 5.25 |
| 21 | 14.83 | 69.91 | 75.41 | 3.83 | 11187 | 45.49 | 21341 | 5.26 |
| 22 | 13.16 | 67.85 | 67.13 | 4.19 | 9926 | 35.81 | 30276 | 4.41 |
| 23 | 15.31 | 67.43 | 71.78 | 2.83 | 10048 | 35.72 | 13493 | 5.16 |
| 24 | 13.80 | 71.97 | 68.02 | 3.17 | 11108 | 58.46 | 23828 | 5.50 |
| 25 | 11.89 | 75.28 | 67.29 | 5.82 | 12845 | 48.88 | 26731 | 5.97 |
| 26 | 19.59 | 62.87 | 68.86 | 5.25 | 8393 | 14.42 | 12410 | 4.26 |
| 27 | 21.21 | 61.00 | 67.31 | 2.41 | 8569 | 17.11 | 13169 | 4.51 |
| 28 | 14.47 | 65.41 | 69.35 | 2.92 | 8536 | 14.82 | 11009 | 5.46 |
| 29 | 20.16 | 65.25 | 71.53 | 1.79 | 8722 | 21.16 | 8312 | 3.58 |
| 32 | 4.10 | 80.89 | 65.94 | 6.79 | 16158 | 63.37 | 1073 | 5.72 |
| 33 | 7.20 | 72.53 | 64.89 | 3.64 | 11796 | 19.85 | 462 | 5.94 |
| 34 | 6.77 | 74.78 | 66.33 | 4.55 | 12931 | 35.85 | 723 | 5.54 |
| 35 | 5.50 | 77.14 | 69.19 | 2.45 | 13155 | 44.91 | 295 | 5.80 |
| 36 | 4.49 | 80.33 | 64.41 | 3.85 | 15616 | 40.24 | 398 | 5.96 |
| 37 | 3.89 | 75.04 | 70.52 | 3.12 | 12466 | 52.56 | 242 | 6.50 |
| 38 | 4.88 | 81.74 | 66.98 | 6.12 | 17157 | 45.94 | 865 | 6.20 |

Lampiran 4. Output data Testing

| | kemiskinan | IPM | TPAK | TPT | PP | PK | SP | PDRB |
|----|------------|-------|-------|------|-------|-------|-------|------|
| 11 | 14.39 | 65.27 | 71.45 | 3.90 | 10429 | 32.28 | 23250 | 5.09 |
| 30 | 7.68 | 77.58 | 65.09 | 3.63 | 11976 | 58.37 | 424 | 5.42 |
| 15 | 5.69 | 79.50 | 64.53 | 4.73 | 14168 | 56.13 | 11909 | 6.05 |
| 31 | 7.44 | 77.58 | 72.21 | 4.06 | 13391 | 40.33 | 616 | 5.83 |

Lampiran 5. Output Pohon Regresi Maksimal

n= 34

node), split, n, deviance, yval
 * denotes terminal node

- 1) root 34 668.01090 11.109120
- 2) SP< 3440 7 10.06109 5.261429 *
- 3) SP>=3440 27 356.52310 12.625190
 - 6) PDRB>=5.055 20 97.87778 11.302500
 - 12) TPT>=3.205 13 32.56152 10.544620 *
 - 13) TPT< 3.205 7 43.98180 12.710000 *
 - 7) PDRB< 5.055 7 123.68420 16.404290 *

Lampiran 6. Output CrossValidation-10 CART

| CP | nsplit | rel error | xerror | xstd |
|----|------------|-----------|-----------|---------------------|
| 1 | 0.45123025 | 0 | 1.0000000 | 1.0146281 0.2359039 |
| 2 | 0.20203433 | 1 | 0.5487697 | 0.9028996 0.1465648 |
| 3 | 0.03193728 | 2 | 0.3467354 | 0.8516809 0.1418459 |
| 4 | 0.01000000 | 3 | 0.3147981 | 0.8516809 0.1418459 |

Lampiran 7. Output Pohon Regresi Optimal

```
Call:
rpart(formula = kemiskinan ~ ., data = kmsTrain, method = "anova",
      control = list(minsplit = 2, maxdepth = 4, cp = 0.01))
n = 34
```

| CP | nsplit | rel error | xerror | xstd |
|----|------------|-----------|-----------|---------------------|
| 1 | 0.45123025 | 0 | 1.0000000 | 1.1614543 0.2735190 |
| 2 | 0.20203433 | 1 | 0.5487697 | 0.9884939 0.1870800 |
| 3 | 0.03193728 | 2 | 0.3467354 | 0.8631222 0.1999623 |
| 4 | 0.01000000 | 3 | 0.3147981 | 0.7847543 0.1647932 |

Variable importance

| IPM | SP | PDRB | PP | TPAK | TPT | PK |
|-----|----|------|----|------|-----|----|
| 20 | 20 | 17 | 16 | 15 | 8 | 4 |

Node number 1: 34 observations, complexity param=0.4512303
mean=11.10912, MSE=19.64738

left son=2 (7 obs) right son=3 (27 obs)

Primary splits:

SP < 3440 to the left, improve=0.4512303, (0 missing)
PDRB < 5.525 to the right, improve=0.4238781, (0 missing)
IPM < 71.98 to the right, improve=0.4223676, (0 missing)
PP < 11782 to the right, improve=0.4197898, (0 missing)
PK < 35.83 to the right, improve=0.2306367, (0 missing)

Surrogate splits:

IPM < 73.845 to the right, agree=0.941, adj=0.714, (0 split)
TPAK < 67.055 to the left, agree=0.941, adj=0.714, (0 split)
PP < 12460 to the right, agree=0.941, adj=0.714, (0 split)
PDRB < 5.895 to the right, agree=0.882, adj=0.429, (0 split)
TPT < 6.115 to the right, agree=0.853, adj=0.286, (0 split)

Node number 2: 7 observations

mean=5.261429, MSE=1.437298

Node number 3: 27 observations, complexity param=0.2020343

mean=12.62519, MSE=13.20456

left son=6 (20 obs) right son=7 (7 obs)

Primary splits:

PDRB < 5.055 to the right, improve=0.3785481, (0 missing)
IPM < 66.19 to the right, improve=0.3579440, (0 missing)

```

PK < 28.175 to the right, improve=0.2642193, (0 missing)
PP < 9245 to the right, improve=0.2432045, (0 missing)
SP < 13875 to the right, improve=0.1987724, (0 missing)
Surrogate splits:
IPM < 65.33 to the right, agree=0.926, adj=0.714, (0 split)
PK < 21.96 to the right, agree=0.852, adj=0.429, (0 split)
PP < 9010.5 to the right, agree=0.815, adj=0.286, (0 split)
TPAK < 67.505 to the right, agree=0.778, adj=0.143, (0 split)
TPT < 2.58 to the right, agree=0.778, adj=0.143, (0 split)

```

```

Node number 6: 20 observations, complexity param=0.03193728
mean=11.3025, MSE=4.893889
left son=12 (13 obs) right son=13 (7 obs)

```

Primary splits:

```

TPT < 3.205 to the right, improve=0.21797030, (0 missing)
IPM < 69.92 to the right, improve=0.16932750, (0 missing)
TPAK < 71.45 to the left, improve=0.13600120, (0 missing)
SP < 14626 to the right, improve=0.08636813, (0 missing)
PK < 33.57 to the left, improve=0.06571339, (0 missing)

```

Surrogate splits:

```

SP < 14626 to the right, agree=0.85, adj=0.571, (0 split)
IPM < 67.37 to the right, agree=0.75, adj=0.286, (0 split)
PP < 8813 to the right, agree=0.75, adj=0.286, (0 split)
PK < 24.865 to the right, agree=0.75, adj=0.286, (0 split)

```

```

Node number 7: 7 observations
mean=16.40429, MSE=17.66917

```

```

Node number 12: 13 observations
mean=10.54462, MSE=2.504733

```

```

Node number 13: 7 observations
mean=12.71, MSE=6.283114

```

Lampiran 8. Output OOB Bagging Regresi

```

Bagging regression trees with 100 bootstrap replications

Call: bagging.data.frame(formula = kemiskinan ~ ., data = persenkemiskinan,
  nbagg = 100, coob = TRUE, conrol = rpart.control(minsplit = 2,
  cp = 0))

Out-of-bag estimate of root mean squared error: 3.1715

```

Lampiran 9. Output CrossValidation-10 Bagging Regresi

```

Bagged CART

34 samples
7 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 31, 31, 31, 30, 30, 32, ...

```

Resampling results:

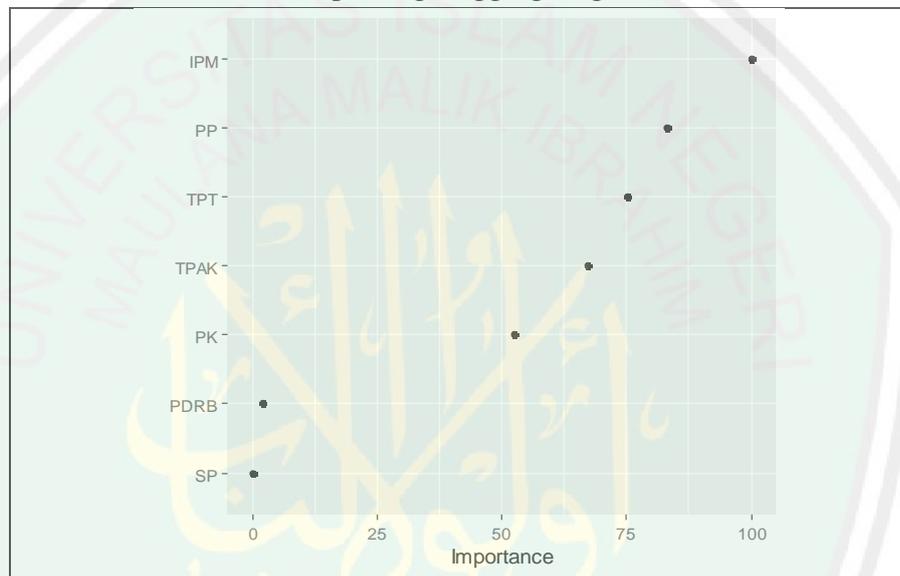
| RMSE | Rsquared | MAE |
|----------|-----------|----------|
| 2.481208 | 0.7012645 | 2.180356 |

Lampiran 10. Output Percobaan Penambahan Pohon pada Plot RMSE

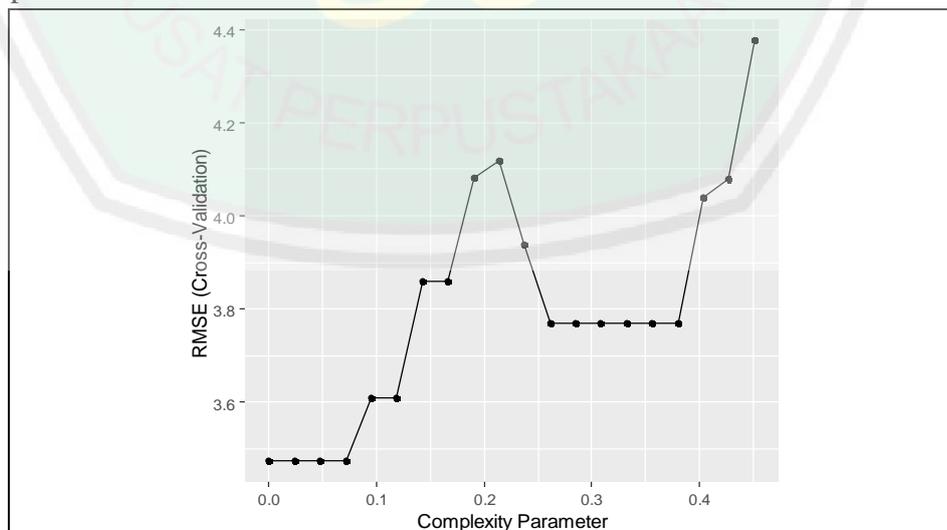
```
> predictions[1:4,1:7]
```

| | result.1 | result.2 | result.3 | result.4 | result.5 | result.6 | result.7 |
|----|----------|----------|----------|----------|----------|----------|----------|
| 25 | 5.50 | 5.50 | 3.89 | 3.89 | 5.69 | 6.77 | 7.68 |
| 6 | 9.56 | 9.56 | 12.02 | 11.42 | 15.31 | 11.42 | 13.80 |
| 2 | 14.83 | 14.83 | 9.98 | 14.83 | 9.72 | 10.37 | 9.72 |
| 33 | 7.27 | 12.11 | 7.80 | 6.77 | 4.49 | 12.11 | 6.77 |

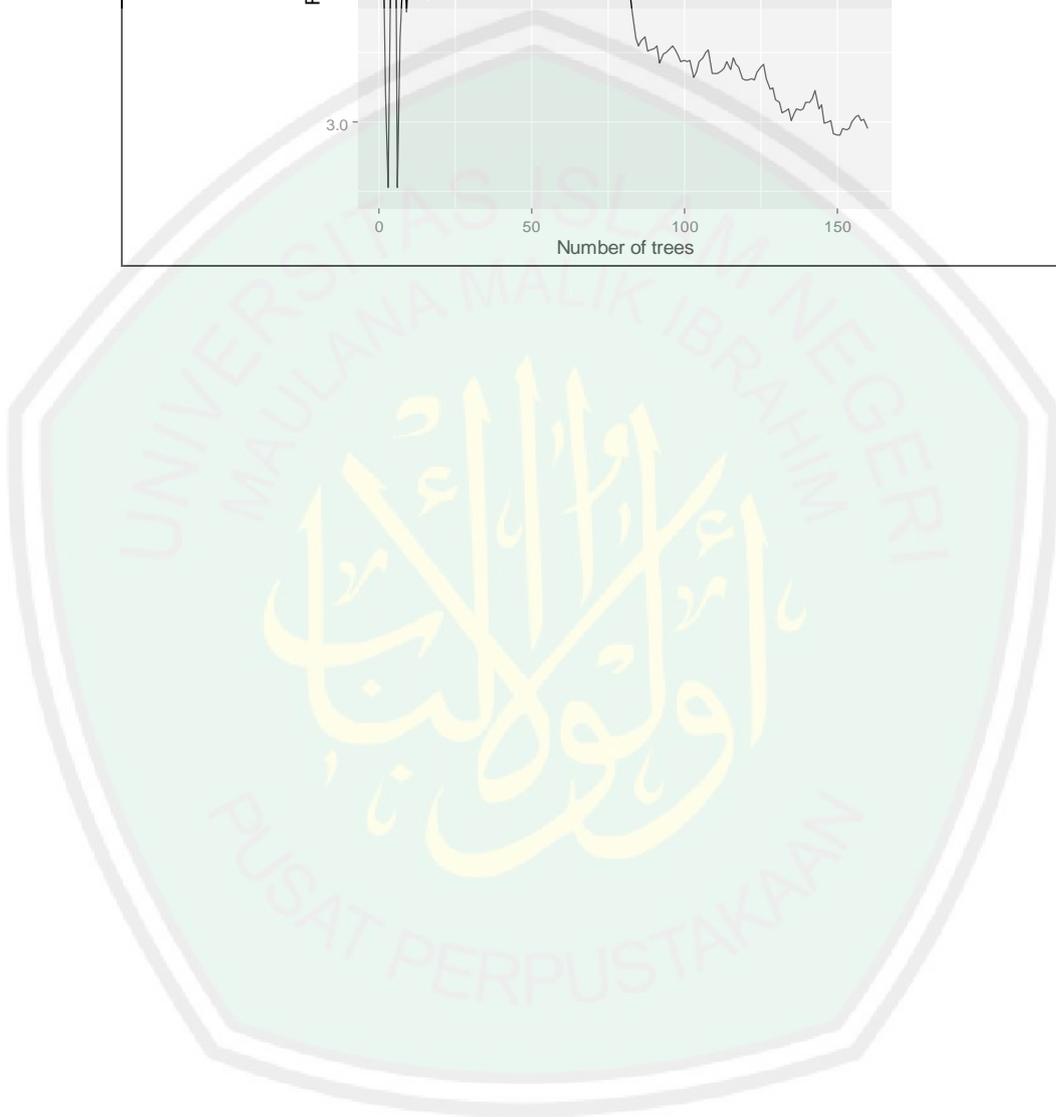
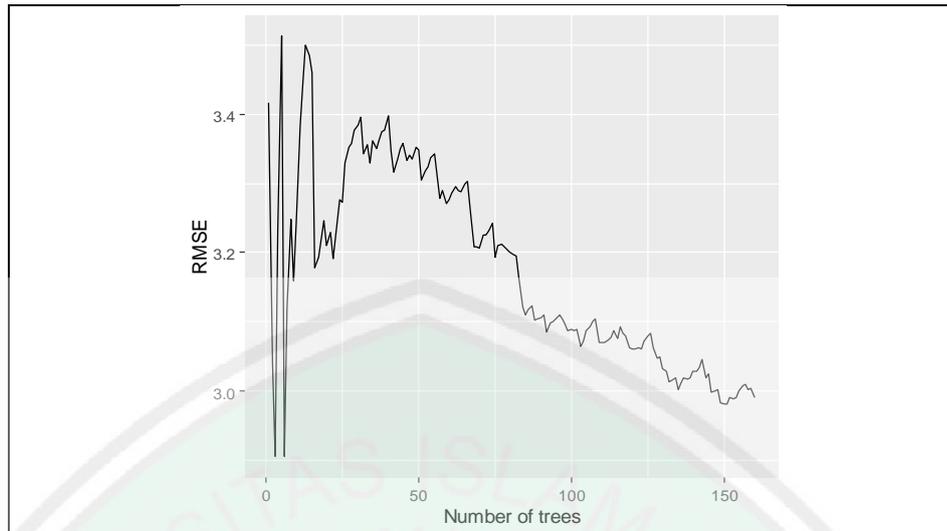
Lampiran 11. Plot Variabel Terpenting Bagging Regresi



Lampiran 12. Plot RMSE CrossValidation CART



Lampiran 13. Plot RMSE Bagging Regresi



RIWAYAT HIDUP



Safrila Shofania, lahir di Malang pada tanggal 28 Juni tahun 1997, akrab dipanggil dengan sebutan Safrila atau Nia. Berdomisili di Jalan Bowling No. 59 Tasikmadu, Kecamatan Lowokwaru, Kelurahan Tasikmadu, Kota Malang. Anak tunggal dari pasangan Bapak Eko Hariadi dan Ibu Luluk Farida. Telah menyelesaikan Pendidikan Taman Kanak-kanak di TK B.A Paramita- Malang dan lulus pada tahun 2004. Kemudian melanjutkan pendidikan Sekolah Dasar di SDN Tasikmadu 1 – Malang dan lulus pada tahun 2010. Selanjutnya melanjutkan pendidikan Sekolah Menengah Pertama di SMPN 16 Malang- Malang dan lulus pada tahun 2013. Pendidikan selanjutnya ditempuh di SMA Laboratorium UM – Malang dan lulus pada tahun 2016 serta pada tahun yang sama melanjutkan kuliah di Universitas Islam Negeri Maulana Malik Ibrahim Malang pada Fakultas Sains dan Teknologi dengan jurusan Matematika murni. Penulis dapat dihubungi melalui email: safilashfnia@gmail.com



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No. 50 Dinoyo Malang Telp./Fax.(0341)558933

BUKTI KONSULTASI SKRIPSI

Nama : Safrila Shofania
 NIM : 16610015
 Fakultas/Jurusan : Sains dan Teknologi/Matematika
 Judul Skripsi : *Pemodelan Kemiskinan Provinsi Jawa Timur Menggunakan Metode Classification And Regression Trees dan Bootstrap Aggregating*
 Pembimbing I : Ria Dhea Layla Nur Karisma, M.Si
 Pembimbing II : Muhammad Khudzaifah, M.Si

| No | Tanggal | Hal | Tanda Tangan |
|-----|------------------|--------------------------------------------------------|--------------|
| 1. | 04 Februari 2020 | Konsultasi Bab I | 1. |
| 2. | 13 Februari 2020 | Konsultasi Bab II | 2. |
| 3. | 20 Februari 2020 | Konsultasi Bab III | 3. |
| 4. | 02 Maret 2020 | Mencari Integrasi ayat Al-Qur'an | 4. |
| 5. | 12 Maret 2020 | Penambahan Integrasi ayat di Bab II | 5. |
| 6. | 30 Maret 2020 | Konsultasi Bab I, II, III, IV dan ACC Seminar Proposal | 6. |
| 7. | 10 Mei 2020 | Perbaikan Integrasi Ayat Al-Qur'an | 7. |
| 8. | 13 Juni 2020 | Konsultasi 1 Persiapan Sidang | 8. |
| 9. | 15 Juni 2020 | Konsultasi 2 Persiapan Sidang | 9. |
| 10. | 16 Juni 2020 | ACC untuk disidangkan | 10. |

Malang, 5 September 2020
 Mengetahui,
 Ketua Jurusan Matematika

Dr. Usman Pagalay, M.Si
 NIP. 19650414 200312 1 001