

IDENTIFIKASI VARIABEL YANG MEMPENGARUHI BESAR PINJAMAN DENGAN METODE POHON REGRESI (Studi Kasus di Unit Pengelola Kegiatan PNPM Mandiri)

Shaumal Luqman¹, Moch. Abdul Mukid², Abdul Hoyyi³

¹Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

^{2,3}Staff Pengajar Jurusan Statistika FSM Universitas Diponegoro
stat.luqman@gmail.com

ABSTRACT

Most people need a loan to fulfil their daily needs, such as a loan of goods or money. Loan can be obtained from financial institutions or individuals. In order to the loan granted by a financial institutions is not wrong target, financial institutions usually apply precaution principle. In making decisions related to how much a decent loan granted to a customer, the financial institutions often use the help of statistical methods. One methods often used is the Classification and Regression Trees (CART). Classification and Regression Trees (CART) is a nonparametric method that can be used to identify the variable that affect the amount of the loan at a financial institutions and estimate how much worth of loans granted. Because of the loan is a continous variable so the form of the tree is a Regression Tree. In this thesis, the financial institutions is UPK PNPM Mandiri Mekar Sejati in Kecamatan Bawang Kabupaten Batang. Variables that may be affected for large loans are age, occupation, type of warranty, the number family members, and the average income per month. The analysis showed that the variables that most influence on the income of the loans. Mean Absolute Percentage Error (MAPE) value from this method is 36%.

Keyword : Regression tree, CART, Large loans.

1. PENDAHULUAN

Undang Undang No. 17 tahun 2007 tentang Rencana Pembangunan Jangka Panjang (RPJP) Nasional tahun 2005-2025 memiliki tujuan salah satunya adalah untuk menjamin tercapainya penggunaan sumber daya secara efisien, efektif, berkeadilan, dan berkelanjutan, serta mengoptimalkan masyarakat. Kemudian dalam PP No. 7 tahun 2005 tentang Rencana Pembangunan Jangka Menengah (RPJM) Nasional 2004-2009 memiliki tiga agenda pembangunan nasional, yaitu menciptakan Indonesia yang aman dan damai, menciptakan Indonesia yang adil dan demokratis, serta meningkatkan kesejahteraan masyarakat.

Menurut tim pengendali PNPM Mandiri (2007), berbagai usaha untuk meningkatkan kesejahteraan masyarakat dengan pendekatan pemberdayaan masyarakat telah dilaksanakan pemerintah melalui berbagai program, diantaranya adalah Program Pengembangan Kecamatan (PPK), yaitu sebagai tahapan pembelajaran dalam pemberdayaan masyarakat. Kemudian PPK dikembangkan lagi menjadi Program Nasional Pemberdayaan Masyarakat (PNPM). PNPM diharapkan dapat menuntaskan tahapan pemberdayaan, yaitu tercapainya kemandirian dan berkelanjutan.

Penyerapan Program Nasional Pemberdayaan Masyarakat Mandiri di Jawa Tengah tahun 2008 mencapai 461 miliar. Sebagian besar dari dana itu digunakan untuk pembangunan sarana fisik. Kepala Badan Pemberdayaan Masyarakat Jawa Tengah,

Kusumardhono menyebutkan, dana terserap di 3.530 desa pada 224 Kecamatan di 29 Kabupaten.

Kabupaten Batang merupakan salah satu Kabupaten yang berada di Jawa Tengah. Pemerintah melaksanakan pemberdayaan masyarakat pada semua wilayah. Namun pada penelitian ini, akan diambil dari salah satu Kecamatan yang berada di Kabupaten Batang. Pemerintah Kabupaten Batang memulai pemberdayaan masyarakat di Kecamatan Bawang pada tahun 2003 sampai 2006 dengan Program Pengembangan Kecamatan (PPK). Kemudian pada tahun 2007 sampai sekarang dilanjutkan dengan Program Nasional Pemberdayaan Mandiri. Kecamatan Bawang merupakan salah satu kecamatan di Kabupaten Batang. PNPM Mandiri Kecamatan Bawang Kabupaten Batang merupakan salah satu upaya untuk mengentaskan kemiskinan dan meningkatkan kesejahteraan masyarakat Kecamatan Bawang Kabupaten Batang. PNPM Mandiri dikelola oleh UPK (Unit Pengelola Kegiatan) kecamatan. UPK berperan sebagai unit pengelola dan operasional pelaksana kegiatan antar desa. UPK mendapatkan penugasan yang diperoleh dari Musyawarah Antar Desa (MAD) atau Badan Kerjasama Antar Desa (BKAD).

Sesuai dengan hasil rapat forum antar Desa Kecamatan Bawang Kabupaten Batang pada bulan Juli tahun 2009 ditetapkan realisasi PNPM Mandiri tahun anggaran 2009 yang diikuti oleh 20 desa di Kecamatan Bawang Kabupaten Batang. Jenis kegiatannya adalah kegiatan pembangunan fisik dan non fisik, kegiatan pembangunan fisik seperti pembangunan sarana dan prasarana, sedangkan kegiatan non fisik seperti simpan pinjam kelompok perempuan.

Koperasi simpan pinjam merupakan perusahaan yang bergerak dalam bidang keuangan. Kegiatan menghimpun dan menyalurkan dana merupakan kegiatan pokok perbankan maupun koperasi simpan pinjam. Fungsi utama bank maupun koperasi simpan pinjam dalam suatu perekonomian adalah untuk memobilisasi dana masyarakat, secara tepat dan cepat menyalurkan dana tersebut kepada pengguna atau investasi yang efektif dan efisien.

Metode regresi berganda dapat dilakukan dengan pendekatan parametrik dan nonparametrik. Dalam pendekatan parametrik terdapat metode yang sering digunakan adalah analisis regresi berganda yang memiliki variabel dependen diharuskan memenuhi asumsi yaitu data harus berdistribusi normal. Dengan adanya keterbatasan metode parametrik, maka digunakan pendekatan nonparametrik. Pada pendekatan nonparametrik terdapat metode dengan berstruktur pohon yaitu metode pohon regresi. Pohon regresi terbentuk jika pada variabel respon memiliki data bertipe kontinu.

Dalam Penelitian ini, akan membahas bagaimana menerapkan bentuk pohon regresi untuk mengidentifikasi variabel-variabel yang mempengaruhi besar pinjaman pada UPK PNPM Mandiri Kecamatan Bawang Kabupaten Batang. Data yang digunakan yaitu variabel dependennya adalah besar pinjaman dan variabel prediktor yaitu umur, jenis pekerjaan, jenis jaminan, banyaknya anggota keluarga dan rata-rata penghasilan per bulan. Data yang digunakan dari UPK PNPM Mandiri Kecamatan Bawang dan Kabupaten Batang.

2. TINJAUAN PUSTAKA

2.1 Metode Pohon Regresi

Pohon regresi adalah metode yang digunakan untuk menggambarkan hubungan antara satu variabel respon yang kontinu dengan satu atau lebih variabel prediktor yang bersifat kontinu dan kategorik. Pohon regresi juga dikenal sebagai pemilahan rekursif secara biner (*binary recursive partitioning*), dimana sekelompok data yang terkumpul dalam suatu ruang yang disebut simpul (*node*) dipilah menjadi dua simpul anak (*child node*), kemudian

setiap simpul anak dapat dipilah lagi menjadi dua simpul anak. Demikian seterusnya, hingga diperoleh suatu simpul akhir (*terminal node*) dan berhenti jika telah memenuhi kriteria tertentu. Adapun variabel yang berpengaruh dalam pohon regresi adalah variabel yang menentukan pemilahan (*splitting*) bagi ruang variabel prediktor.

Pohon regresi dibentuk melalui suatu pemilahan rekursif berdasarkan aturan pemilahan tersebut. Proses pemilahan dilakukan pada tiap simpul untuk membentuk pohon regresi dengan aturan sebagai berikut:

1. Mencari semua kemungkinan pemilah pada tiap variabel prediktor berdasarkan aturan pemilahan tersebut.
2. Memilih pemilah yang terbaik pada masing-masing variabel prediktor, kemudian melakukan pemilihan dari kumpulan pemilah terbaik tersebut.

Pemilah terbaik (*goodness of split*) adalah pemilah yang memaksimumkan ukuran kehomogenan di dalam masing-masing simpul anak relatif terhadap simpul induknya dan yang memaksimumkan ukuran pemisahan (*separation*) antara kedua simpul anak tersebut. Pemilah terbaik dihitung berdasarkan selisih jumlah kuadrat deviasi (*total squared deviation*) antara simpul induk dengan simpul anak pemilahannya. Selisih terbesar akan dijadikan sebagai pemilah terbaik. Misalkan diketahui simpul t berisi amatan-amatan $\{(x_{1(t)}, y_{1(t)}), (x_{2(t)}, y_{2(t)}), (x_{3(t)}, y_{3(t)}), \dots, (x_{n(t)}, y_{n(t)})\}$, sedangkan $n(t)$ adalah jumlah amatan dalam simpul t maka rata-rata respon dalam simpul t adalah (Breiman, *et al.*, 1984).

$$\bar{y}(t) = \frac{1}{n(t)} \sum_{i=1}^n y_{i(t)}$$

sehingga rata-rata jumlah kuadrat deviasi yang digunakan sebagai kriteria kehomogenan pada suatu simpul t adalah

$$R(t) = \frac{1}{N} \sum_{i=1}^n (y_{i(t)} - \bar{y}(t))^2$$

Jika terdapat pemilah s yang memilah t menjadi simpul anak kiri t_L dan simpul anak kanan t_R , maka kriteria selisih jumlah kuadrat deviasi adalah

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

dengan pemilah terbaik s^* dari t adalah pemilahan pada S yang sedemikian hingga

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

2.2 Penghentian Ukuran Pohon Regresi Optimal

Pohon regresi yang terbentuk berukuran besar dan kompleks dalam menggambarkan data. Sehingga perlu dilakukan suatu pemangkasan. Pemangkasan dilakukan dengan memangkas bagian pohon yang kurang penting sehingga didapat pohon yang optimal. Metode pemangkasan pohon sebelumnya menghasilkan urutan *subtree* $T_1 > T_2 > \dots > \{t_1\}$. Karena pohon regresi yang dihasilkan begitu banyak, maka pemilihan pohon yang optimal perlu dilakukan. Pohon regresi yang optimal selanjutnya dapat digunakan untuk kepentingan prediksi. Namun demikian pohon regresi ini juga memiliki tingkat kesalahan prediksi.

Ada beberapa cara yang digunakan untuk menduga tingkat kesalahan prediksi dari satu model pohon regresi (Breiman, *et al.*, 1984).

a. Penduga Penggantian (*Resubstitution Estimate*)

Penduga penggantian adalah rata-rata kuadrat kesalahan dari keseluruhan data sebagai ukuran tingkat kesalahan prediksi. Penduga penggantian ini dirumuskan dengan

$$R(T_k) = \frac{1}{N} \sum_{(x_n, y_n) \in \mathcal{E}} (y_n - d_k(x_n))^2$$

b. Penduga Sampel Uji (Test Sample Estimate)

Amatan dibagi dua secara acak menjadi *training sample* \mathcal{E}_1 dan *test sample* \mathcal{E}_2 . \mathcal{E}_1 digunakan untuk membentuk deretan pohon regresi $\{T_k\}$ dan memperoleh nilai penduga $d_k(x)$ setelah melalui proses pemangkasan, dan \mathcal{E}_2 digunakan untuk memperoleh $R^{ts}(T_k)$. Jika \mathcal{E}_2 berukuran N_2 dan penduga yang sesuai untuk pohon T_k dinotasikan dengan $d_k(x_n)$, maka penduga sampel uji adalah

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

Pohon regresi optimal adalah T_{k0} yang memenuhi kriteria

$$R^{ts}(T_{k0}) = \min_k (R^{ts}T_k)$$

2.3 Prediksi dalam Pohon Regresi

Misalkan akan mencari nilai prediksi dari y_i yang berkaitan dengan vektor input \mathbf{x}_i . Apabila pohon regresi yang optimal sudah diperoleh, dengan memperhatikan nilai-nilai vektor \mathbf{x}_i maka amatan ke- i dapat ditelusuri sedemikian sehingga ia masuk ke simpul terminal tertentu dari pohon regresi tersebut. Selanjutnya nilai \hat{y}_i ditentukan sama dengan nilai rata-rata dari simpul terminal dimana amatan ke- i tersebut masuk.

2.4 Ketelitian Dugaan Pohon Regresi

Ketelitian dugaan pohon regresi digunakan untuk mengetahui besar pengaruh hasil prediksi terhadap amatannya pohon regresi. Menurut Makridakis, *et al.*, (1999), salah satu ukuran yang digunakan untuk melihat nilai kesalahan persentase dengan menggunakan kriteria *Mean Absolute Percentage Error (MAPE)*.

$$MAPE = \frac{1}{N} \sum_{i=1}^n \left| \frac{x_t - f_t}{x_t} \right|$$

3. METODE PENELITIAN

Data yang digunakan pada kasus di atas adalah data besar pinjaman pada UPK PNPM Mandiri Kecamatan Bawang Kabupaten Batang. Dengan variabel respon besar pinjaman dan variabel prediktor berjumlah 5 variabel yaitu:

- Y = Besar pinjaman
- X₁ = Umur
- X₂ = Jenis pekerjaan
- X₃ = Jenis jaminan
- X₄ = Banyaknya anggota keluarga
- X₅ = Rata-rata penghasilan per bulan

Langkah-langkah yang akan dilakukan untuk mencapai penelitian ini adalah analisis deskriptif dan pohon regresi yang dapat dijelaskan sebagai berikut:

- a. Untuk mengetahui karakteristik besar pinjaman di UPK PNPM Mandiri Kecamatan Bawang Kabupaten Batang akan dilakukan analisis deskriptif dengan menggunakan software SPSS 20.

b. Mencari bentuk pohon regresi terhadap besar pinjaman di UPK PNPM Mandiri dengan metode CART (*Classification and Regression Tree*). Analisis menggunakan *software* SPSS 20 dan dilakukan dengan tahapan sebagai berikut:

Eksplorasi data yaitu pembagian data menjadi dua yaitu data *learning* dan data *testing*. Data *learning* digunakan untuk proses pembentukan pohon regresi, sedangkan data *testing* digunakan untuk melihat kemampuan pohon regresi yang terbentuk dalam menduga atau memprediksi data baru.

Pembentukan pohon regresi yaitu dengan tahapan berikut:

1. Pertumbuhan (*growing*) pohon regresi maksimal, diawali dengan pemilahan yang dilakukan dengan cara memilah data hasil pengujian besar pinjaman sesuai aturan pertumbuhan (*growing*) dan aturan pemilahan. Pemilahan terbaik diantara variabel predictor ditentukan berdasarkan selisih terbesar jumlah kuadrat deviasi antara simpul induk dan kedua simpul anak yang memilahnya.
2. Penghentian pemilahan pohon regresi akan berhenti apabila sudah tidak dimungkinkan lagi dilakukan pemilahan. Hal ini dikarenakan hanya terdapat satu amatan yang ada dalam simpul akhir atau adanya batasan minimum.
3. Menentukan pohon regresi yang optimal.

4. HASIL DAN PEMBAHASAN

Hasil eksplorasi data besar pinjaman dari 218 peminjam, menunjukkan bahwa besarnya rata-rata besar pinjaman adalah sebesar 3087155,96. Nilai minimumnya adalah sebesar 1000000 dan nilai maksimumnya adalah sebesar 5000000. Simpangan baku dari besar pinjaman adalah sebesar 1406607,247.

Tabel 1 Statistik Deskriptif Besar Pinjaman

Minimum	Maksimum	Rata-rata	Standardevisi
1000000	5000000	3087155,96	1406607,247

Jumlah data keseluruhan yang digunakan dalam penelitian ini adalah 218. Data penelitian dibagi menjadi data *learning* (80% dengan jumlah data 174) dan data *testing* (20% dengan jumlah data 44). Data *learning* digunakan untuk proses pembentukan pohon regresi, sedangkan data *testing* digunakan untuk menguji ketepatan model.

1. Proses Pertumbuhan (*Growing*) Pohon Regresi Maksimal

Pembentukan pohon regresi dilakukan menurut aturan pemilahan dan aturan pertumbuhan, yaitu dimulai dari pemilahan data peminjaman dari 174 peminjam oleh variabel pemilah terbaik. Tiap-tiap variabel prediktor memiliki jumlah kemungkinan pemilah yang berbeda, yaitu:

Untuk variabel umur terdapat sebanyak 15 nilai amatan yang berbeda, maka akan terdapat pemilah sebanyak $15 - 1 = 14$ kemungkinan.

Untuk variabel pekerjaan terdapat sebanyak 4 nilai amatan yang berbeda, maka akan terdapat pemilah sebanyak $2^{4-1} - 1 = 7$ kemungkinan.

Untuk variabel jaminan terdapat sebanyak 4 nilai amatan yang berbeda, maka akan terdapat pemilah sebanyak $2^{4-1} - 1 = 7$ kemungkinan.

Untuk variabel banyaknya anggota keluarga terdapat sebanyak 4 nilai amatan yang berbeda, maka akan terdapat pemilah sebanyak $6 - 1 = 5$ kemungkinan.

Untuk variabel penghasilan terdapat sebanyak 10 nilai amatan yang berbeda, maka akan terdapat pemilah sebanyak $10 - 1 = 9$ kemungkinan.

Sebelum dilakukan proses pemilahan simpul, terlebih dahulu memilih variabel terbaik dari kelima variabel prediktor. Pemilahan variabel terbaik berdasarkan kriteria

goodness of split. Pemilah terbaik yang terpilih untuk simpul 0 (simpul akar) adalah variabel penghasilan. Variabel penghasilan terpilih karena memiliki nilai $\Delta R(s,t) = R(t) - R(t_L) - R(t_R)$.

Berikut adalah ilustrasi perhitungan nilai pemilah terbaik pada simpul akar untuk variabel penghasilan ≤ 1100000 dan penghasilan > 1100000 :

$$R(t) = \frac{1}{174} ((5000000 - 3132184)^2 + (3000000 - 3132184)^2 + (5000000 - 3132184)^2 + \dots + (2000000 - 3132184)^2)$$

$$= 1942297529396$$

$$R(t_L) = \frac{1}{174} ((3000000 - 2368932)^2 + (4000000 - 2368932)^2 + (4000000 - 2368932)^2 + \dots + (2000000 - 2368932)^2)$$

$$= 896440129450$$

$$R(t_R) = \frac{1}{174} ((5000000 - 4239437)^2 + (5000000 - 4239437)^2 + (3000000 - 4239437)^2 + \dots + (4000000 - 4239437)^2)$$

$$= 200744698073$$

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

$$= 1942297529396 - 896440129450 - 200744698073$$

$$= 845112701873$$

Setelah terbentuk dan terpilih pemilah terbaik, maka simpul akar yang terdiri dari 174 data dipilah menjadi 2 simpul. Simpul 1 terbentuk akibat criteria variable penghasilan $\leq 1.100.000$ dan simpul 2 terbentuk akibat criteria variable penghasilan $> 1.100.000$.

2. Proses Penghentian Pemilahan Simpul

Proses pemilahan simpul yang berulang-ulang akan berhenti ketika $N(t) \leq 5$ atau semua nilai y pada suatu simpul adalah sama. Selanjutnya proses tersebut menghasilkan pohon maksimal.

3. Pohon Regresi Optimal

Setelah terbentuk pohon regresi maksimal, selanjutnya akan dilakukan pemilihan satu buah pohon regresi optimal. Dalam mencari pohon regresi yang optimal digunakan kriteria penduga penggantian (*resubstitution estimate*). Nilai $R(T_k)$ untuk berbagai kemungkinan nilai k ada di Tabel 2. Berdasarkan Tabel 2, diketahui bahwa setelah $k = 4$, penurunan nilai $R(T_k)$ tidak lagi sebesar nilai-nilai k sebelumnya. Sehingga ditetapkan pohon regresi yang optimal adalah pohon regresi dengan 4 simpul terminal.

Tabel 2 Nilai $R(T_k)$

Banyak Terminal Node (k)	$R(T_k)$	$R(T_k) - T(T_{k+1})$
1	1942297529396	845112701873
2	1097184827523	207915064251
3	889269763272	101549045122
4	787720718150	8696459898
5	779024258252	34745484401
6	744278773852	10541670031
7	733737103820	6345368257
8	727391735563	6385696041
9	721006039523	6170695140
10	714835344382	

Selanjutnya akan dibahas mengenai pemilihan pohon regresi optimal dengan menggunakan kriteria penduga sampel uji (*test sampel estimate*). Nilai $R^{ts}(T_k)$ untuk berbagai nilai k dapat dilihat di Tabel 3.

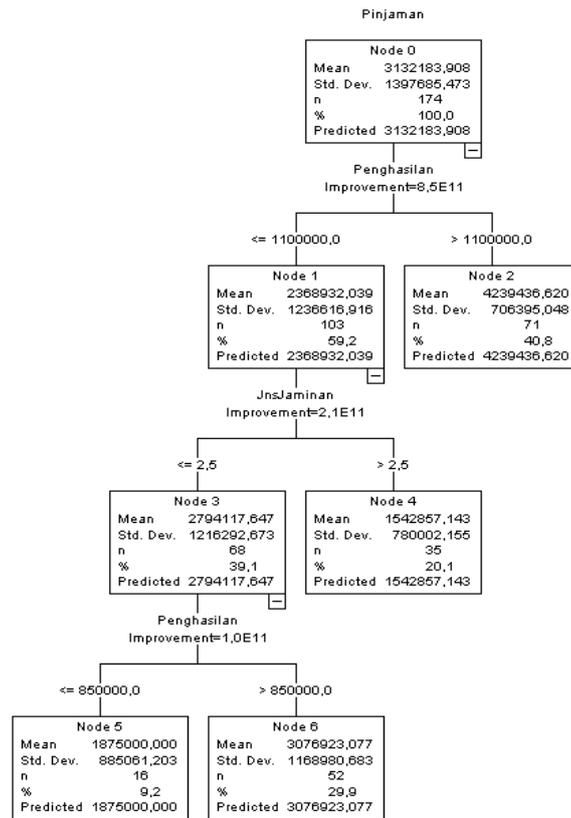
Tabel 3 Nilai $R^{ts}(T_k)$

Banyak Terminal Node (k)	$R^{ts}(T_k)$
1	2086960609856
2	998015004034
3	1000084536761
4	709084351274
5	724153076609
6	721353818724
7	739000396445
8	727851106981
9	722800601930
10	722776162257

Pada Tabel 3 terlihat bahwa dari 10 *subtree* yang terbentuk, *subtree* ke-4 yang terpilih. Hal ini dikarenakan nilai *subtree* ke-4 memiliki nilai penduga kehomogenan $R^{ts}(T_k)$ yang lebih kecil dari pada *subtree* yang lainnya.

Pohon regresi optimal yang terpilih tersebut terlihat pada Gambar dibawah ini memiliki 3 simpul dalam (*internal node*) dan 4 simpul terminal (*terminal node*).

Dari Gambar dibawah ini dapat disimpulkan bahwa variabel penghasilan dan jenis jaminan memiliki pengaruh paling penting dalam melakukan pinjaman pada UPK PNPM Mandiri Kecamatan Bawang.



4. Ketelitian Dugaan Pohon Regresi

Pohon regresi yang baik adalah pohon yang mampu memprediksi data atau amatan baru. Untuk tujuan tersebut maka sebanyak 44 data *testing* dimasukkan kedalam pohon regresi optimal yang terbentuk untuk melakukan pengukuran ketelitian dugaan, yaitu melihat kemampuan pohon regresi yang terbentuk dalam menduga atau memprediksi besar pinjaman untuk data *testing*. Kriteria yang digunakan adalah *Mean Absolute Percentage Error (MAPE)*

$$\begin{aligned}
 MAPE &= \frac{1}{N} \sum_{i=1}^n \left| \frac{x_i - f_i}{x_i} \right| \\
 &= \frac{1}{44} \sum_{i=1}^{44} \left(\left(\frac{5000000 - 4239437}{5000000} \right) + \left(\frac{5000000 - 4239437}{5000000} \right) + \left(\frac{4000000 - 3076923}{4000000} \right) + \dots + \left(\frac{1000000 - 1542857}{1000000} \right) \right) \\
 &= 36\%
 \end{aligned}$$

5. KESIMPULAN

Berdasarkan hasil dari pembahasan dapat disimpulkan bahwa dalam pinjaman di UPK PNPM Mandiri Mekar Sejati memperoleh variabel penting yang pertama dalam melakukan pinjaman yaitu variabel penghasilan. Selanjutnya pengaruh yang paling penting kedua adalah variabel jenis jaminan. Berdasarkan kasus tersebut diketahui besar pengaruh hasil prediksi terhadap amatannya dengan menggunakan ketelitian dugaan pohon regresi.

Ukuran yang digunakan untuk mengetahui nilai kesalahan persentase pada pohon regresi dengan melihat nilai *Mean Absolute Percentage Error (MAPE)* yaitu sebesar 36%.

6. DAFTAR PUSTAKA

- [1] Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984, *Classification And Regression Trees*, Chapman & Hall, New York.
- [2] Makridakis, S., Wheelwright, C, S., McGee, E, V. 1999. *Metode dan aplikasi peramalan*, Edisi 2. Binarupa Aksara, Jakarta.
- [3] Tim Pengendali PNPM Mandiri. 2007. *Pedoman Umum PNPM Mandiri*. Jakarta.