

Klasifikasi Minat Korban Banjir terhadap Pembelian Produk Asuransi Dampak Banjir Menggunakan *Software Orange* (Studi Kasus: Kota Jakarta Timur)

Ira Rosianal Hikmah^{1*}, Yulial Hikmah²

Program Studi Rekayasa Keamanan Siber, Politeknik Siber dan Sandi Negara¹, Program Pendidikan Vokasi, Universitas Indonesia²

*Email: ira.rosianal@poltekssn.ac.id

Info Artikel

Kata Kunci:

banjir, asuransi, *supervised learning*, *software orange*, evaluasi model, kurva ROC

Keywords:

flood, *insurance*, *supervised learning*, *software orange*, *model evaluation*, *ROC curve*

Tanggal Artikel

Dikirim: 08 Oktober 2021

Direvisi: 18 Oktober 2021

Diterima: 30 November 2021

Abstrak

Hampir semua kota di Indonesia mengalami banjir setiap tahun, termasuk DKI Jakarta. Berdasarkan data Badan Nasional Penanggulangan Bencana (BNPB) tahun 2020, Kota Jakarta Timur merupakan kota yang rawan banjir. Banjir merupakan bencana yang relatif paling banyak menimbulkan kerugian. Kerugian yang diakibatkan oleh banjir, terutama kerugian tidak langsung, dapat menempati urutan pertama atau kedua setelah gempa bumi atau tsunami. Menurut BNPB, Kota Jakarta Timur terkena dampak banjir terparah saat bencana banjir awal tahun 2020. Oleh karena itu, perlu adanya upaya mitigasi bencana untuk meminimalisir kemungkinan terjadinya risiko banjir. Salah satu mitigasi risiko akibat bencana banjir adalah dengan membeli produk asuransi dampak banjir sebagai upaya pemindahan risiko yang mungkin akan terjadi. Namun, tidak semua orang membeli produk asuransi dampak banjir karena faktor ekonomi dan sosial. Penelitian ini bertujuan untuk melakukan prediksi terhadap minat penduduk Kota Jakarta Timur terutama yang mengalami banjir terhadap pembelian produk asuransi dampak banjir. Oleh karena itu, penelitian ini menggunakan algoritma *supervised learning* dengan tujuh pilihan model dan diperoleh tiga model yang dapat direkomendasikan adalah model *Random Forest*, *Naive Bayes*, dan *SVM* yang diperoleh berdasarkan kriteria evaluasi yaitu waktu yang dibutuhkan untuk *training* dan *testing*, tingkat akurasi, dan presisi, serta kurva ROC.

Abstract

*Almost all cities in Indonesia experience flooding every year, including DKI Jakarta. Based on data from the National Disaster Management Agency (BNPB) in 2020, East Jakarta City is a city that is prone to flooding. Flooding is a disaster that relatively causes the most losses. Losses caused by floods, especially indirect losses, can rank first or second after an earthquake or tsunami. According to BNPB, East Jakarta City was worst affected by flooding during the early 2020 flood disaster. Therefore, disaster mitigation efforts are needed to minimize the possibility of flood risk. One of the risk mitigations due to floods is to buy flood impact insurance products to transfer risks that may occur. However, not everyone buys flood impact insurance products due to economic and social factors. This study aims to predict the interest of residents of East Jakarta City, especially those who experience flooding, to purchase flood impact insurance products. Therefore, this study used a supervised learning algorithm with seven model choices and obtained three models that can be recommended, namely the *Random Forest*, *Naive Bayes*, and *SVM* models obtained based on the evaluation criteria, namely the time required for training and testing, the level of accuracy, and precision, as well as the ROC curve.*

1. PENDAHULUAN

Posisi Wilayah Negara Kesatuan Republik Indonesia yang terletak di garis khatulistiwa dan berbentuk kepulauan, menimbulkan potensi yang tinggi terhadap berbagai jenis bencana yang berkaitan dengan hidrometeorologi, seperti banjir, banjir bandang, kekeringan, cuaca ekstrim (tornado), gelombang dan abrasi yang kuat serta kebakaran hutan dan lahan. Fenomena perubahan iklim juga meningkatkan ancaman bencana hidrometeorologi [1]. Banjir merupakan ancaman bencana dengan risiko tinggi di Indonesia khususnya terhadap harta benda dan infrastruktur serta merupakan ancaman serius bagi perekonomian masyarakat. Banjir dapat disebabkan oleh kondisi alam yang statis (seperti geografi, topografi, dan geometri alur sungai), peristiwa alam yang dinamis (seperti curah hujan yang tinggi, pembendungan dari laut/pasang surut pada sungai utama, penurunan tanah dan berbentuk segitiga akibat sedimentasi), dan aktivitas-aktivitas manusia yang dinamis seperti penggunaan lahan dataran banjir yang tidak tepat (membangun pemukiman di bantaran sungai, kurangnya infrastruktur pengendalian banjir, penurunan permukaan tanah dan kenaikan permukaan laut akibat pemanasan global) [1]. Secara umum, dampak banjir dapat bersifat langsung maupun tidak langsung. Dampak langsung relatif lebih mudah diprediksi dibandingkan dampak tidak langsung [2].

Berdasarkan Kajian Risiko Bencana Banjir, terdapat sepuluh provinsi prioritas yang menjadi fokus kegiatan penanggulangan bencana banjir, salah satunya adalah Provinsi DKI Jakarta [1]. Secara geomorfologi, sebagian besar wilayah Jakarta terbentuk oleh proses fluvial dan terletak di bagian utara Pulau Jawa. Kondisi ini menyebabkan wilayah DKI Jakarta rawan bencana hidrometeorologi, khususnya banjir. Berdasarkan kondisi tersebut menunjukkan bahwa Provinsi DKI Jakarta merupakan daerah yang memiliki ancaman banjir yang tinggi [3]. Bencana banjir yang terjadi di awal tahun 2020 merupakan salah satu banjir terbesar di Jakarta [4]. Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) mencatat hujan ekstrem yang terjadi merata di Jakarta dan sekitarnya pada 31 Desember 2019, dan 1 Januari 2020. Curah hujan tertinggi mencapai 377 mm per hari, lebih tinggi dari rekor curah hujan tertinggi pada Februari 2015 yaitu 367 mm per hari (Kompas, 3 Januari 2020) [5]. Kondisi ini diperparah dengan letak Jakarta yang merupakan dataran rendah yang rawan banjir. Menurut Sakethi (2010), luas wilayah DKI Jakarta yang berada di bawah permukaan laut adalah 40% dari total luas daratan [6], [7].

Banjir juga merupakan bencana yang relatif paling banyak menimbulkan kerugian. Kerugian akibat banjir, terutama kerugian tidak langsung, dapat menempati urutan pertama atau kedua setelah gempa atau tsunami [2]. Kerusakan akibat banjir dapat dibagi menjadi empat jenis: kerusakan berwujud langsung (misalnya, kerusakan fisik akibat kontak dengan air), tidak langsung (misalnya, hilangnya produksi dan pendapatan), langsung tidak berwujud (misalnya, hilangnya nyawa), dan tidak langsung (misalnya, trauma) [8]. Kerugian akibat banjir dapat berupa kerusakan bangunan, kehilangan barang berharga, dan kerugian yang mengakibatkan tidak masuk kerja dan sekolah. Banjir tidak dapat dicegah, tetapi dapat dikendalikan, dan dampak kerugian dapat dikurangi [9]. Oleh karena itu, upaya mitigasi dan pengendalian banjir harus dilakukan untuk mengurangi risiko bencana guna meminimalkan kerugian material dan korban jiwa yang mungkin terjadi. Menurut BNPB dalam Buku Pegangan Tanggap Bencana (2017), salah satu langkah masyarakat pra bencana dalam menghadapi bencana banjir adalah dengan mempertimbangkan pembelian produk asuransi dampak banjir [10].

Penelitian ini bertujuan untuk membuat model yang dapat mengklasifikasikan minat masyarakat Indonesia khususnya korban banjir untuk membeli produk asuransi bencana banjir berdasarkan faktor-faktor sosial dan ekonomi serta faktor pemahaman terhadap asuransi dan produk asuransi dampak banjir. Penelitian ini melakukan *data mining* dengan menggunakan algoritma *supervised learning* dengan teknik klasifikasi dan regresi. Klasifikasi merupakan analisis data prediktif dimana data terbagi menjadi beberapa kelas yang sudah diketahui dan diberikan label sedangkan regresi merupakan proses melihat hubungan dan pengaruh pada suatu dataset, yang terbagi menjadi satu variabel dependen dan satu atau lebih variabel independen [11]. Ada beberapa strategi untuk mengetahui seberapa baik model yang terbentuk. Salah satunya adalah validasi silang. Metode validasi silang yang paling sederhana adalah metode *holdout* dimana data dibagi menjadi dua bagian yaitu data *training* untuk menemukan parameter model, dan data testing untuk menguji keakuratan parameter atau model yang terbentuk [12]. Beberapa tahun yang lalu, terdapat banyak perangkat lunak *data mining* yang telah dikembangkan. Beberapa dari perangkat lunak tersebut tersedia secara bebas dan gratis [13]. Perangkat lunak *open source* merupakan perangkat lunak komputer dimana *source code* tersedia untuk umum. Pengguna dapat menggunakan, memeriksa, memperbaharui, dan mendistribusikan kepada siapa saja untuk digunakan. Salah satu perangkat lunak *data mining open source* adalah Orange yang dapat digunakan untuk visualisasi data dan analisis data [13]. Orange menyediakan pemodelan baik supervised maupun unsupervised learning. Bahkan dengan *widget test and score*, pengguna dapat menjalankan beberapa model sekaligus pada perangkat lunak tersebut. Orange juga menyediakan fitur evaluasi model

seperti akurasi, presisi, waktu yang dibutuhkan untuk training dan testing, spesifisitas, dan sebagainya [14]–[17]. Penelitian ini melakukan pengambilan sampel terhadap korban banjir tahun 2021 yang bertempat tinggal di Jakarta Timur dengan targetnya adalah menentukan apakah berminat atau tidak untuk membeli produk asuransi dampak banjir. Dengan data tersebut, dilakukan klasifikasi *data mining* dengan beberapa algoritma *supervised learning* dan menggunakan perangkat lunak *open-source* Orange. Penelitian ini melakukan evaluasi pada model yang terbentuk dengan menghasilkan tingkat akurasi, tingkat presisi, waktu training, dan waktu testing, serta menampilkan kurva ROC.

2. METODE PENELITIAN

2.1 Data Penelitian

Populasi dalam penelitian ini adalah penduduk Indonesia. Penelitian ini menggunakan teknik *non-probability purposive sampling*, yaitu teknik pengambilan sampel karena adanya pertimbangan tertentu [18]. Berdasarkan laporan BNPB (Badan Nasional Penanggulangan Bencana), Kota Jakarta Timur terkena dampak banjir terparah di awal tahun 2020 [19], [20]. Selanjutnya, Kepala Pusat Data, Informasi, dan Komunikasi Kebencanaan BNPB, Agus Wibowo mengatakan bahwa delapan kecamatan di Jakarta Timur dengan jumlah korban banjir terbanyak, yakni 752 Kepala Keluarga (2476 jiwa) [21], [22]. Karena keterbatasan waktu, biaya, dan sumber daya, penelitian ini memilih beberapa Rukun Tetangga yang pernah dan masih mengalami banjir pada awal tahun 2020. Sebaran responden terpilih dapat dilihat pada Tabel 1.

Tabel 1. Sebaran data sampel penelitian

| <i>Kecamatan</i> | <i>Kabupaten</i> | <i>Jumlah Sampel</i> |
|------------------|------------------|----------------------|
| Duren Sawit | Pondok Bambu | 10 |
| | Bidara Cina | 25 |
| Jatinegara | Kampung Melayu | 22 |
| | Rawa Bunga | 10 |
| Kramat Jati | Cawang | 25 |
| | Cililitan | 10 |
| Makasar | Cipinang Melayu | 18 |
| Pulo Gadung | Pisangan Timur | 20 |
| Total | | 140 |

Jumlah sampel penelitian sebanyak 140 keluarga. Pengumpulan data dilakukan melalui wawancara dan juga dengan mengisi kuesioner online. Alasan dilakukannya wawancara adalah karena beberapa pertanyaan harus diberikan secara langsung untuk melihat kebenaran dan keabsahan jawaban responden. Namun, meski dilakukan dengan wawancara, penelitian ini tetap mematuhi protokol kesehatan karena pandemi covid-19. Proses pengumpulan data melalui wawancara dan pengisian kuesioner secara daring dapat dilihat pada Gambar 1 dan Gambar 2. Data penelitian terdiri atas 29 atribut yang dapat dilihat pada Tabel 2.



Gambar 1. Tampilan Kuesioner online



Gambar 2. Dokumentasi Survei

Tabel 2. Atribut penelitian

| No | Target dan Atribut | Kode |
|----|---|------------|
| 1 | Minat membeli produk asuransi dampak banjir (Tidak Berminat, Berminat) | Y (Target) |
| 2 | Status pernikahan (Belum Menikah, Menikah, Cerai) | X1 |
| 3 | Status kepemilikan rumah (Kontrak, Milik Keluarga, Milik Pribadi) | X2 |
| 4 | Usia kepala rumah tangga | X3 |
| 5 | Tingkat Pendidikan kepala rumah tangga (Tidak Bersekolah, SD, SMP, SMA, D3, S1, S2/S3) | X4 |
| 6 | Pendapatan bulanan (0, 0-2 juta, 2-4 juta, 4-6 juta, 6-8 juta, 8-10 juta, 10-12 juta, >12 juta) | X5 |
| 7 | Rata-rata biaya bulanan | X6 |
| 8 | Jumlah tanggungan | X7 |
| 9 | Pengalaman banjir (Tidak Pernah, Pernah) | X8 |
| 10 | Tahun lalu (2019) mengalami banjir (Tidak, Ya) | X9 |
| 11 | Berapa kali pernah mengalami banjir? | X10 |
| 12 | Tinggi banjir (cm) | X11 |
| 13 | Rumah dekat sungai (Tidak, Ya) | X12 |
| 14 | Berapa waktu yang dibutuhkan hingga banjir surut? (Hari) | X13 |
| 15 | Pernah mengungsi (Tidak Pernah, Pernah) | X14 |
| 16 | Berapa kali pernah mengungsi setiap tahunnya? | X15 |
| 17 | Jika terjadi banjir, saya selalu mengungsi (Tidak, Ya) | X16 |
| 18 | Anggota keluarga pernah ada yang meninggal karena banjir (Tidak Pernah, Pernah) | X17 |
| 19 | Banjir menyebabkan kerusakan properti saya (Tidak, Ya) | X18 |
| 20 | Berapa kerugian kerusakan properti karena banjir? (Rupiah) | X19 |
| 21 | Saya tahu apa itu asuransi (Tidak Setuju, Setuju) | X20 |
| 22 | Saya setuju dengan pernyataan berikut: Asuransi adalah perjanjian antara tertanggung dan penanggung yang mewajibkan tertanggung membayar sejumlah premi untuk memberikan penggantian atas risiko yang akan terjadi (Tidak Setuju, Setuju) | X21 |
| 23 | Saya setuju dengan pernyataan berikut: Fungsi asuransi adalah pengendalian risiko yang akan terjadi (Tidak Setuju, Setuju) | X22 |
| 24 | Saya setuju dengan pernyataan berikut: Asuransi merupakan salah satu cara meminimalisir kerugian dari kehilangan atau kerusakan benda-benda berharga yang terjadi akibat Banjir (Tidak Setuju, Setuju) | X23 |
| 25 | Saya tahu jenis-jenis asuransi (Tidak Setuju, Setuju) | X24 |
| 26 | Saya tahu beberapa perusahaan asuransi di Indonesia (Tidak Setuju, Setuju) | X25 |

| No | Target dan Atribut | Kode |
|----|--|------|
| 27 | Saya tahu tujuan asuransi (Tidak Setuju, Setuju) | X26 |
| 28 | Saya tahu bahwa ada produk asuransi yang dapat dibeli untuk menjamin kerugian yang akan Anda terima akibat terkena banjir setiap tahunnya (Tidak Setuju, Setuju) | X27 |
| 29 | Saya tahu cara membeli produk asuransi (Tidak Setuju, Setuju) | X28 |
| 30 | Keluarga saya ada yang mengikuti asuransi (Tidak Setuju, Setuju) | X29 |

2.2 Model Supervised Learning dan Evaluasi Model

Terdapat empat *tools* yang digunakan untuk *Machine Learning*, yaitu *Supervised Learning*, *Reinforcement Learning*, *Active Learning*, atau *Unsupervised Learning*. Dua *tools* yang umum digunakan adalah *Supervised* dan *Unsupervised Learning* [23]. Teknik *Data Mining* dapat dikelompokkan menjadi tiga tipe yaitu klasifikasi, regresi, dan *clustering* [11]. Penelitian ini menggunakan *tool Supervised Learning* dengan teknik klasifikasi dan regresi. Klasifikasi *Data Mining* adalah penempatan objek-objek ke salah satu dari beberapa kelas yang telah ditentukan sebelumnya. Klasifikasi banyak digunakan untuk memprediksi kelas pada suatu label tertentu dengan membangun model berdasarkan data *training* dan menggunakan model tersebut untuk mengklasifikasikan data *testing* [24]. Pada penelitian ini, terdapat 2 kelas, yaitu kelas 1 berarti tidak berminat dan kelas 2 menunjukkan berminat membeli produk asuransi dampak banjir. Regresi merupakan teknik *Data Mining* yang termasuk ke dalam *supervised learning* dan digunakan untuk memprediksi target numerik seperti regresi linier sederhana [25]. Namun jika variabel dependennya merupakan variabel kategori, maka Regresi Logistik dapat digunakan. Jika variabel dependennya hanya dua kategori maka disebut Regresi Logistik binomial namun jika lebih dari dua kategori maka disebut dengan Regresi Logistik multinomial [26]. Terdapat 7 (tujuh) metode yang digunakan pada penelitian ini, yaitu metode Decision Tree, Random Forest, Support Vector Machine (SVM), Neural Network, Naïve Bayes, k-NN, dan Regresi Logistik.

Setelah model terbentuk, dilakukan evaluasi. Ukuran yang digunakan dalam penelitian ini adalah tingkat akurasi dan presisi serta waktu *training* dan *testing*. Formula untuk tingkat akurasi dapat dihitung berdasarkan hasil *confusion matrix* sebagai berikut: [13]

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Selanjutnya, presisi merupakan nilai prediksi positif, atau dapat dihitung dengan formula berikut: [13]

$$Presisi = \frac{TP}{TP+FP} \quad (2)$$

dengan *True Positive* (TP) dan *True Negative* (TN) memberikan informasi ketika *classifier* dalam melakukan klasifikasi data bernilai benar sedangkan *False Positive* (FP) dan *False Negative* (FN) memberikan informasi ketika *classifier* salah dalam melakukan klasifikasi data [27].

2.3 Software Orange

Orange merupakan perangkat lunak *open source* untuk *Machine Learning* dan *Data Mining* yang ditulis dengan Bahasa Python. Orange dikembangkan oleh Laboratorium Bioinformatika, Fakultas Ilmu Komputer dan Informasi, Universitas Ljubljana [28]. *Software Orange* memiliki beberapa *widget* yang dapat digunakan. Penelitian ini menggunakan *widget test and score* untuk melakukan pemodelan, prediksi, dan evaluasi model. Selain itu, digunakan *widget* untuk *input* data dan *processing* data seperti *widget file*, *data table*, dan beberapa model dalam penelitian ini yaitu *Tree*, *Random Forest*, *Neural Network*, *Naïve Bayes*, kNN, SVM, dan *Logistic Regression*. Evaluasi lanjutan juga dilakukan dengan menggunakan *widget ROC analysis*.

2.4 Langkah-Langkah Penelitian

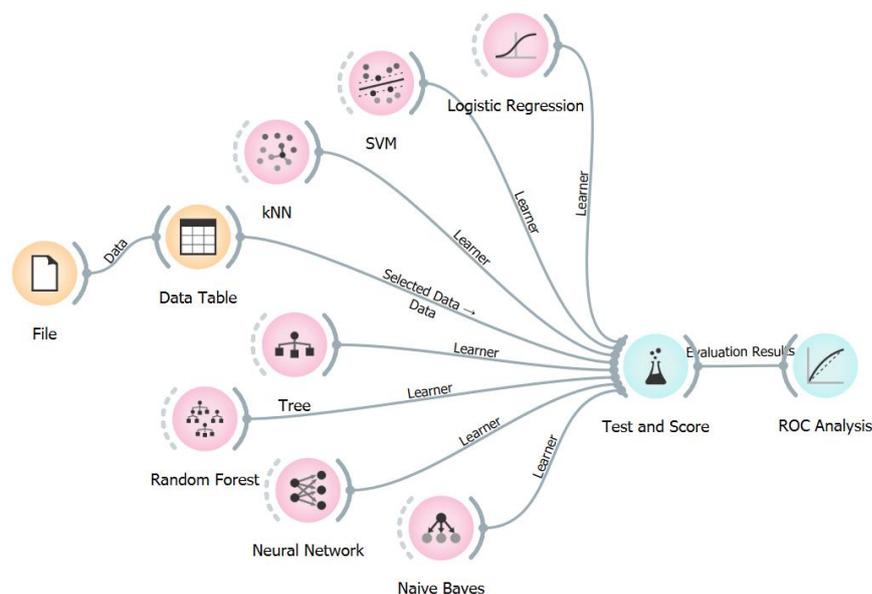
Secara umum, proses klasifikasi data mencakup dua langkah. Langkah pertama adalah membangun model atau aturan klasifikasi dengan mempelajari data *training* dengan label kelas terkait. Langkah kedua adalah menggunakan

model tersebut untuk melakukan klasifikasi terhadap data *testing* untuk mendapatkan keakuratan dari model atau aturan klasifikasi tersebut [28]. Langkah-langkah secara rinci dalam penelitian ini adalah sebagai berikut:

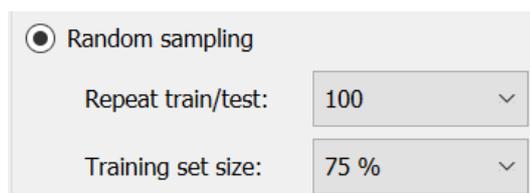
1. *Data cleaning*. Pada tahap ini, dilakukan pembersihan data untuk menghindari hal-hal seperti *incomplete*, *noisy*, dan *inconsistent*.
2. *Data integration*. Pada tahap ini, dilakukan perubahan atau konversi data ke jenis data yang diinginkan.
3. Menentukan target data
4. *Preprocessing*. Tahap ini dilakukan sebelum proses *Data Mining*. Pada tahap ini dilakukan pengambilan sampel (*sampling*) secara acak untuk membaginya menjadi data *training* dan data *testing* dengan rasio 3:1. Penelitian ini melakukan pengulangan *sampling* sebanyak 100 kali.
5. Proses *Data Mining*.
6. Evaluasi model.

3. HASIL DAN PEMBAHASAN

Workflow penelitian ini dengan *software* Orange dapat dilihat pada Gambar 3. Pada Gambar 3, *workflow* penelitian ini dapat dibagi menjadi tiga bagian, yaitu tahap persiapan data yang berwarna kuning, model / *learner* penelitian yang berwarna merah, dan prediksi serta evaluasi yang berwarna hijau. Pada bagian sebelumnya, telah dituliskan bahwa penelitian ini melakukan pengulangan pengambilan sampel sebanyak 100 kali dengan rasio data *training* dan data *testing* yaitu 3:1. Dengan demikian, dilakukan konfigurasi pada *widget test and score* dengan memilih 100 pada opsi *repeat train/test* serta 75% pada opsi *training set size*. Tampilan konfigurasi *widget test and score* dapat dilihat pada Gambar 4.



Gambar 3. *Workflow* Penelitian



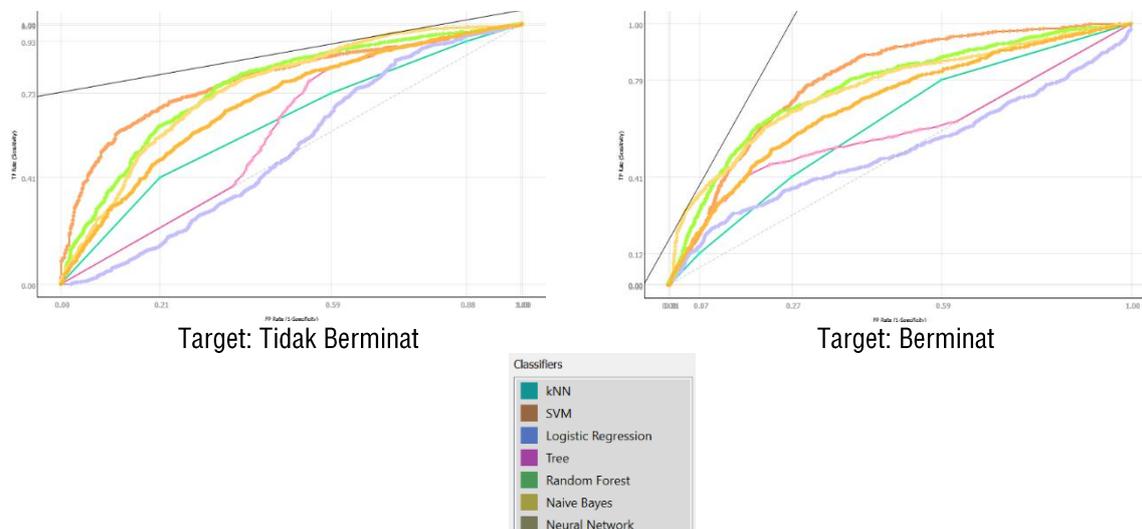
Gambar 4. Konfigurasi *Widget Test and Score*

Hasil evaluasi ketujuh model penelitian ini dapat dilihat pada Gambar 5. Terlihat bahwa berdasarkan kriteria evaluasi penelitian ini, yaitu waktu yang dibutuhkan untuk *training* dan *testing*, model *Neural Network* membutuhkan waktu yang paling lama.

| Evaluation Results | | | | |
|---------------------|----------------|---------------|-------|-----------|
| Model | Train time [s] | Test time [s] | CA | Precision |
| kNN | 2.052 | 1.186 | 0.747 | 0.682 |
| Logistic Regression | 2.332 | 1.091 | 0.738 | 0.700 |
| Naive Bayes | 0.790 | 0.157 | 0.736 | 0.775 |
| Neural Network | 25.070 | 1.669 | 0.753 | 0.732 |
| Random Forest | 3.275 | 1.158 | 0.783 | 0.749 |
| SVM | 4.795 | 1.510 | 0.762 | 0.708 |
| Tree | 3.834 | 0.007 | 0.744 | 0.728 |

Gambar 5. Hasil Evaluasi Model *Supervised Learning*

Kurva ROC ketujuh model penelitian ini dapat dilihat pada Gambar 6. Terlihat bahwa model terbaik dalam penelitian ini, yaitu *Random Forest*, *Naïve Bayes*, dan *SVM*, baik untuk memprediksi tidak berminat atau berminat membeli produk asuransi dampak banjir. Ketiga model di atas yang paling mewakili tingkat akurasi dan presisi terbaik berdasarkan Gambar 5. Model yang tidak direkomendasikan dalam penelitian ini adalah Regresi Logistik, Tree, dan kNN. Model *Neural Network* termasuk ke dalam *moderate evaluation*. Namun, jika melihat efisiensi waktu yang ditampilkan pada Gambar 5, maka model tersebut juga kurang direkomendasikan untuk data penelitian ini.



Gambar 6. Kurva ROC ketujuh model

4. KESIMPULAN

Berdasarkan proses dan hasil *data mining*, terlihat bahwa *software* Orange merupakan salah satu perangkat lunak yang mudah digunakan dan mempercepat proses *data mining*. Berdasarkan hasil penelitian, data penelitian dapat direkomendasikan untuk memprediksi minat masyarakat khususnya korban banjir, untuk membeli produk asuransi dampak banjir dengan menggunakan pemodelan *supervised learning* yaitu *Random Forest*, *Naive Bayes*, atau *SVM*. Ketiga model tersebut masing-masing menghasilkan tingkat akurasi dan (presisi) adalah 78,3% (74,9%), 73,6% (77,5%), dan 76,2% (70,8%). Model *Neural Network* bisa menjadi pilihan keempat dalam penelitian ini, tetapi waktu yang dibutuhkan untuk *training* sangat lama dibandingkan dengan model lain.

DAFTAR PUSTAKA

- [1] BNPB RI, "Rencana Nasional Penanggulangan Bencana 2015-2019," 2014. [Daring]. Tersedia pada: https://www.bnpb.go.id/uploads/renas/1/BUKU_RENAS_PB.pdf.
- [2] A. Rosyidie, "Banjir: Fakta dan Dampaknya, serta Pengaruh dari Perubahan Guna Lahan," *J. Reg. City Plan.*, vol. 24, no. 3, hal. 241–249, 2013, doi: <https://doi.org/10.5614/jpwk.2013.24.3.1>.
- [3] S. Dahlia, N. H. Tricahyono, dan W. F. Rosyidin, "Analisis Kerawanan Banjir Menggunakan Pendekatan Geomorfologi di DKI Jakarta," *J. Alami J. Teknol. Reduksi Risiko Bencana*, vol. 2, no. 1, hal. 1–8, 2018.
- [4] E. Y. Gunawibawa dan H. Oktiani, "Politik & Bencana Banjir Jakarta 2020: Analisis Peta Percakapan #JakartaBanjir," *Expo. J. Ilmu Komun.*, vol. 3, no. 1, hal. 60–75, 2020, doi: <https://doi.org/10.33021/exp.v3i1.989>.
- [5] A. M. Ginting, "Dampak Ekonomi dan Kebijakan Mitigasi Risiko Banjir di DKI Jakarta dan Sekitarnya Tahun 2020," *Info Singkat*, vol. XII, no. 1, hal. 19–24, 2020.
- [6] BPPTPDAS Surakarta, "Kajian Banjir Jakarta 1 Januari 2020," Surakarta, 2020.
- [7] T. M. Sakethi, *Mengapa Jakarta Banjir*. Jakarta: PT. Mirai Sakethi, 2010.
- [8] P. Sidi, *Pemanfaatan Ilmu Aktuaria dalam Mewujudkan Jaminan Risiko Banjir di dalam Konsep Smart City*. Universitas Terbuka (UT), 2017.
- [9] A. Findayani, "Kesiap Siagaan Masyarakat dalam Penanggulangan Banjir di Kota Semarang," *J. Geogr. Media Inf. Pengemb. dan Profesi Kegeografian*, vol. 2, no. 1, hal. 102–114, 2015, doi: <https://doi.org/10.15294/jg.v12i1.8019>.
- [10] BNPB RI, "Buku Saku Tanggap Tangkas Tangguh Menghadapi Bencana," 2017. [Daring]. Tersedia pada: https://siaga.bnpb.go.id/hkb/po-content/uploads/documents/Buku_Saku-10Jan18_FA.pdf.
- [11] M. Roos, "A data analysis demonstrator for managing customer experience in a partnering ventur," Stellenbosch University, 2019.
- [12] T. Wendler dan S. Grottrup, *Data Mining with SPSS Modeler: TheORY, Exercises and Solution*, 2nd ed. Switzerland: Springer Nature Switzerland, 2021.
- [13] R. Ratra dan P. Gulia, "Experimental Evaluation of Open Source Data Mining Tools (WEKA and Orange)," *Int. J. Eng. Trends Technol.*, vol. 68, no. 8, hal. 30–35, 2020.
- [14] J. Demsar dan B. Zupan, "Orange: Data mining Fruitful and Fun," *Informatica*, vol. 37, hal. 55–60, 2013.
- [15] Orange, "Orange Data Mining: Fruitful and Fun." [Daring]. Tersedia pada: <https://orangedatamining.com/>.
- [16] Orange Data Mining, "Orange Data Mining Library Documentation Release 3."
- [17] B. Zupan dan Dems, "Introduction to data mining," 2018. doi: 10.1007/978-3-642-19721-5_1.
- [18] Sugiyono, *Memahami Penelitian Kualitatif*. Bandung: Alfabeta, 2014.
- [19] BNPB RI, "Hujan Lebat Sebabkan 23 Kecamatan DKI Jakarta Terdampak Banjir," 2020. <https://bnpb.go.id/berita/hujan-lebat-sebabkan-23-kecamatan-dki-jakarta-terdampak-banjir> (diakses Jun 01, 2020).
- [20] K. Nisa, "Rekapitulasi Data Banjir DKI Jakarta dan Penanggulangannya Tahun 2020," 2020. <https://statistik.jakarta.go.id/rekapitulasi-data-banjir-dki-jakarta-dan-penanggulangannya-tahun-2020/> (diakses Jun 01, 2020).
- [21] D. Andayani, "BNPB: Hujan Sebabkan 23 Kecamatan di DKI Jakarta Terdampak Banjir," *detikNews*, Feb 08, 2020. <https://news.detik.com/berita/d-4891589/bnpb-hujan-sebabkan-23-kecamatan-didki-jakarta-terdampak-banjir> (diakses Jun 01, 2020).
- [22] CNN Indonesia, "BNPB Sebut Banjir Rendam 23 Kecamatan di DKI, Jaktim Terparah," Feb 08, 2020. <https://www.cnnindonesia.com/nasional/20200208215738-20-472935/bnpb-sebut-banjir-rendam-23-kecamatan-di-dki-jaktim-terparah> (diakses Jun 01, 2020).
- [23] S. Shalev-Shwartz dan S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. New York: Cambridge University Press, 2014.
- [24] P. Meilina, "Penerapan Data Mining dengan Metode Klasifikasi Menggunakan Decision Tree dan Regresi," *J. Teknol. Univ. Muhammadiyah Jakarta*, vol. 7, no. 1, hal. 11–20, 2015.
- [25] S. Sansgiry, M. Bhosle, dan K. Sail, "Factors That Affect Academic Performance Among Pharmacy Students," *Am. J. Pharm. Educ.*, vol. 70, no. 5, hal. Article 104, 2006.
- [26] H. Zhou, *Learn Data Mining Through Excel: A Step-by-Step Approach for Understanding Machine Learning Methods*. USA: Apress, 2020.

- [27] J. Han dan M. Kamber, *Data Mining: Concepts and Techniques Tutorial*. San Francisco: Morgan Kaufman Publisher, 2001.
- [28] A. Naik dan L. Samant, "Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime," *Procedia Comput. Sci.*, vol. 85, no. 2016, hal. 662–668, 2016, doi: 10.1016/j.procs.2016.05.251.