

**PREDIKSI JUMLAH RILIS *GAME* PADA PLATFORM
STEAM DENGAN MENGGUNAKAN ALGORITMA
*K-NEAREST NEIGHBOR REGRESSION***

SKRIPSI

**OLEH:
SAYID SYEH YUSUF AL-BALGAIST
NIM. 16610112**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**PREDIKSI JUMLAH RILIS *GAME* PADA PLATFORM
STEAM DENGAN MENGGUNAKAN ALGORITMA
*K-NEAREST NEIGHBOR REGRESSION***

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
Sayid Syeh Yusuf Al-Balgaist
NIM. 16610112**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**PREDIKSI JUMLAH RILIS *GAME* PADA PLATFORM
STEAM DENGAN MENGGUNAKAN ALGORITMA
*K-NEAREST NEIGHBOR REGRESSION***

SKRIPSI

**Oleh
Sayid Syeh Yusuf Al-Balgaist
NIM. 16610112**

Telah Disetujui Untuk Diuji
Malang, 21 Juni 2023

Dosen Pembimbing I,



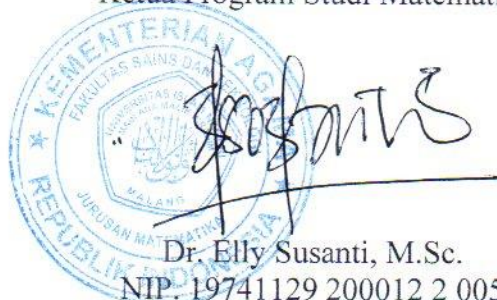
Hisyam Fahmi, M.Kom.
NIP. 198907272019031018

Dosen Pembimbing II,



Mohammad Nafie Jauhari, M.Si.
NIDT. 19870218201608011056

Mengetahui,
Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc.
NIP. 19741129 200012 2 005

**PREDIKSI JUMLAH RILIS *GAME* PADA PLATFORM
STEAM DENGAN MENGGUNAKAN ALGORITMA
*K-NEAREST NEIGHBOR REGRESSION***

SKRIPSI

Oleh
Sayid Syeh Yusuf Al-Balgaist
NIM. 16610112

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Menemperoleh Gelar Sarjana Matematika (S.Mat)
Tanggal 21 Juni 2023


Ketua Penguji : Juhari, M.Si.

Anggota Penguji 1 : Muhammad Khudzaifah, M.Si.

Anggota Penguji 2 : Hisyam Fahmi, M. Kom.

Anggota Penguji 3 : Mohammad Nafie Jauhari, M.Si.

Mengetahui,
Ketua Program Studi Matematika


Dr. Elly Susanti, M.Sc.
NIP. 19741129 200012 2 005

PERNYATAAN KEASLIAN TULISAN

Saya yang bertandatangan di bawah ini:

Nama : Sayid Syeh Yusuf Al-Balgaist
NIM : 16610112
Program Studi : Matematika
Fakultas : Sains dan Teknologi
Judul Skripsi : Prediksi Jumlah Rilis *Game* Pada Platform Steam Dengan Menggunakan Algoritma *K-Nearest Neighbor Regression*

menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilan 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 di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 26 Juni 2023

Yang membuat pernyataan,



Sayid Syeh Yusuf Al-Balgaist
NIM. 16610112

MOTO

“Jika kamu percaya dengan impianmu aku akan membuktikan padamu bahwa kamu bisa meraih impianmu hanya dengan bekerja keras.”

PERSEMBAHAN

Bismillahirrahmanirrahim, segala puji bagi Allah Swt. dengan ketulusan hati,
skripsi ini peneliti persembahkan kepada:

Kedua orang tua, alm bapak Sayyid Jafar dan ibu Lilis Rinayati tercinta yang
senantiasa memberikan dukungan dan doa yang tiada hentinya demi keberhasilan
penulis.

Teruntuk kepada paman (Jamal Abdul Naser) yang juga senantiasa memberikan
dukungan baik moril maupun materil saat proses mengerjakan skripsi ini.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah memberikan rahmat, hidayah serta karunia-Nya. Tak lupa shalawat serta salam selalu tercurah kepada Baginda Nabi Muhammad SAW, yang telah membimbing umat manusia menuju jalan kebenaran. Berkat taufiq dan hidayah-Nya penulis dapat menyelesaikan penyusunan skripsi ini dengan judul “**Prediksi Jumlah Rilis Game pada Platform Steam dengan Menggunakan Algoritma *K-Nearest Neighbor Regression***” sebagai salah satu syarat untuk memperoleh gelar sarjana dalam bidang matematika di Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Sepanjang proses penyusunan skripsi ini, penulis banyak mendapatkan bimbingan, arahan, dukungan, dan semangat dari beberapa pihak. Oleh karena itu, penulis sampaikan ucapan terimakasih sebesar-besarnya kepada:

1. Bapak Prof. Dr. H. M. Zainuddin, MA. selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Ibu Dr. Sri Harini, M.Si. selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Ibu Dr. Elly Susanti, M.Sc. selaku Ketua Program Studi Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Bapak Hisyam Fahmi, M.Kom, selaku dosen pembimbing penulis yang telah memberikan banyak waktu, bimbingan, serta arahan selama perkuliahan hingga dapat terselesaikannya skripsi.
5. Bapak Mohammad Nafie Jauhari, M.Si, selaku dosen pembimbing Integrasi Sains dan Islam yang telah banyak meluangkan waktunya untuk membimbing penulis tentang sains dan perspektif agama islam.
6. Ibu Evawati Alisah, M.Pd selaku dosen wali matematika yang telah memberikan banyak dukungan, bimbingan dan arahan kepada penulis.
7. Seluruh dosen Program Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

8. Orang tua dan seluruh keluarga besar yang selalu senantiasa mendoakan, memberikan dukungan, semangat, serta kasih sayang sehingga peneliti dapat menyelesaikan tugas akhir.
9. Teman-teman angkatan 2016 dan teman-teman penulis yang telah memberikan semangat dan motivasi kepada penulis agar segera menyelesaikan penulisan skripsi ini.

Dengan mengharapkan rida dari Allah SWT, semoga amal baik dari semua pihak telah disebutkan diterima di sisi Allah. Penulis juga berharap semoga skripsi ini bermanfaat kepada orang yang membacanya.

Wassalamu 'alaikum Warahmatullahi Wabarakatuh

Malang,2023

Peneliti

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTO	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR SIMBOL	xiv
DAFTAR LAMPIRAN	xv
ABSTRAK	xvi
ABSTRACT	xvii
مستخلص البحث	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah.....	5
1.6 Definisi Istilah.....	5
BAB II KAJIAN TEORI	7
2.1 <i>Data Time Series</i>	7
2.2 Regresi.....	8
2.3 <i>Machine Learning</i>	9
2.3.1 <i>Algoritma K-Nearest Neighbor Regression</i>	10
2.4 Perhitungan Jarak.....	11
2.4.1 <i>Euclidean Distance</i>	11
2.5 <i>K-Fold Cross Validation</i>	12
2.6 <i>Loss Function</i>	13
2.6.1 <i>MAD</i>	13
2.7 Kajian Integrasi Topik dengan Islam	14
2.8 Kajian Integrasi Topik dengan Teori Pendukung	14
2.9 Steam.....	16
BAB III METODE PENELITIAN	18
3.1 Pendekatan Penelitian	18
3.2 Data dan Sumber Data	19
3.3 Teknik Analisis Data.....	20
3.4 <i>Flowchart</i> Perhitungan Prediksi.....	23
BAB IV PEMBAHASAN.....	24
4.1 Persiapan Data.....	24
4.2 Transformasi Data.....	25

4.3	Pemrosesan Data	27
4.4	Analisa Metode KNN <i>Regression</i>	33
4.5	Prediksi dalam Islam	38
BAB V PENUTUP		40
5.1	Kesimpulan	40
5.2	Saran.....	40
DAFTAR PUSTAKA		42
LAMPIRAN.....		44
RIWAYAT HIDUP		46

DAFTAR TABEL

Tabel 3.1 <i>Sample</i> Data Steam dari Kaggle.com.....	19
Tabel 3.2 Data <i>Time Series</i> pada $p = 14$	19
Tabel 4.1 Jumlah Data Semua <i>Game</i> yang Rilis Perhari	24
Tabel 4.2 Data Semua <i>Game</i> dalam Bentuk <i>Time Series</i> dengan 7 Variabel ..	25
Tabel 4.3 Jumlah Data Berdasarkan Jumlah Variabel Input.....	26
Tabel 4.4 Urutan Terkecil Jarak Data <i>Test</i> ke-1	29
Tabel 4.5 Perbandingan Nilai Aktual dan Prediksi	30
Tabel 4.6 Nilai MAD pada Setiap <i>Fold</i>	31
Tabel 4.7 Nilai MAD pada Setiap Data	33
Tabel 4.8 Perbandingan Nilai MAD pada Setiap Data	38

DAFTAR GAMBAR

Gambar 2.1 Model Regresi	8
Gambar 2.2 <i>K-Fold Cross Validation</i>	12
Gambar 3.1 <i>Flowchart</i> Perhitungan Prediksi	23
Gambar 4.1 Diagram Rata-rata Nilai MAD pada Setiap k	32
Gambar 4.2 Diagram Rata-rata Nilai MAD pada Setiap Parameter	33
Gambar 4.3 Plot Perbandingan Hasil Prediksi Data Semua <i>Game</i>	34
Gambar 4.4 Plot Perbandingan Hasil Prediksi Data Aksi	35
Gambar 4.5 Plot Perbandingan Hasil Prediksi Data Petualangan	35
Gambar 4.6 Plot Perbandingan Hasil Prediksi Data Kasual	36
Gambar 4.7 Plot Perbandingan Hasil Prediksi Data Simulasi	36
Gambar 4.8 Plot Perbandingan Hasil Prediksi Data Strategi	37
Gambar 4.9 Plot Perbandingan Hasil Prediksi Data RPG	37

DAFTAR SIMBOL

i	: Iterasi, untuk $i = 1, 2, \dots, n$
k	: k dari KNN <i>Regression</i>
p	: Panjang variabel input
w	: Bobot
n	: Jumlah partisi <i>K-Fold</i>
d_{euclid}	: Jarak <i>Euclidean</i>
$x_{test(i)}$: Data <i>testing</i> ke- i
$x_{train(i)}$: Data <i>training</i> ke- i
$y_{test(i)}$: Data target <i>testing</i> ke- i
$y_{train(i)}$: Data target <i>training</i> ke- i
X	: Jumlah seluruh data
X_{test}	: Jumlah data <i>test</i>
X_{train}	: Jumlah data <i>train</i>

DAFTAR LAMPIRAN

Lampiran 1 Data Steam.....	43
Lampiran 2 <i>Source Code</i>	43

ABSTRAK

Al-Balgaist, Sayid Syeh Yusuf. 2023. **Prediksi Jumlah Rilis *Game* pada Platform Steam dengan Menggunakan Algoritma *K-Nearest Neighbor Regression***. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Hisyam Fahmi, M.Kom. (II) Mohammad Nafie Jauhari, M.Si.

Kata Kunci: KNN *Regression*, Prediksi, *Game*, MAD dan Steam

Peningkatan jumlah *game* yang rilis pada platform Steam disebabkan karna adanya peningkatan kemajuan teknologi yang pesat. Para pengembang khususnya pengembang *indie games* mengalami kesulitan dalam menentukan waktu rilis karena harus bersaing dengan *game* lain yang rilis bersamaan. Tujuan peneliti adalah memprediksi jumlah rilis *game* pada platform Steam dengan menggunakan *K-Nearest Neighbor Regression*. Data yang dipakai sebanyak 1826 merupakan data sekunder dari tahun 2018 hingga 2022 yang diunduh dari laman Kaggle.com dan disusun menjadi data *time series*. Metode yang dilakukan yaitu persiapan data, transformasi data, pemrosesan data dan evaluasi metode yang digunakan. Dengan menerapkan pencarian parameter optimal yaitu penggunaan jumlah variabel input dari 1 hingga 14 variabel, dan KNN *Regression* dari $k=3$ hingga $k=20$ menghasilkan nilai MAD pada tujuh jenis data yaitu: data semua *game* memiliki nilai 5,64 pada $k=19$ dan delapan variabel input, data aksi memiliki nilai 2,34; data petualangan 2,46; data kasual 2,23; data simulasi 1,60 dan data RPG 1,64 di mana ke lima data tersebut memiliki parameter yang sama pada $k=20$ dan 14 variabel input, dan data strategi 1,60 pada $k=19$ dan 14 variabel input. Berdasarkan nilai MAD terkecil yang diperoleh dapat disimpulkan bahwa penelitian ini dapat digunakan untuk memprediksi jumlah *game* yang rilis pada platform Steam.

ABSTRACT

Al-Balgaist, Sayid Syeh Yusuf. 2023. **The Prediction of Game Release Number on Steam Platform Using Algorithms K-Nearest Neighbor Regression.** Thesis. Department of Mathematics, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisor: (I) Hisyam Fahmi, M.Kom. (II) Mohammad Nafie Jauhari, M.Si.

Keywords: KNN Regression, Prediction, Game, MAD and Steam

The increase in the number of games released on the Steam platform is due to the rapid increase in technological advancements. Developers, especially indie game developers, have difficulty determining the release time because they have to compete with other games released simultaneously. The researcher's goal was to predict the number of game releases on the Steam platform using K-Nearest Neighbor Regression. The data used as many as 1826 is secondary data from 2018 to 2022 downloaded from the Kaggle.com page and compiled into time series data. The methods carried out are data preparation, data transformation, data processing, and evaluation of the methods used. Applying optimal parameter search, namely the use of the number of input variables from 1 to 14 variables, and KNN Regression from $k = 3$ to $k = 20$ produces MAD values in seven types of data as follows: the data of all games has a value of 5,64 at $k=19$ and eight input variables, action's data has a value of 2,34; Adventure Data 2,46; casual data 2,23; simulation data 1,60 and RPG data 1,64 where the five data have the same parameters at $k = 20$ and 14 input variables, and strategy data 1,60 at $k = 19$ and 14 input variables. Based on the smallest MAD obtained value, it can be concluded that this study can be used to predict the number of games released on the Steam platform.

مستخلص البحث

البقيس ، سيد شيخ يوسف. ٢٠٢٣. تنبؤ عدد الألعاب التي تم إصدارها على منصة Steam باستخدام خوارزمية الانحدار *K-Nearest Neighbor*. البحث العلمي. قسم الرياضيات. كلية العلوم والتكنولوجيا ، جامعة مولانا مالك إبراهيم الإسلامية الحكومية بمالانج. المشرف : (١) هشام فهمي، الماجستير. (٢) محمد نافع جوهرى ، الماجستير.

الكلمات المفتاحية: انحدار KNN ، تنبؤ ، ألعاب ، MAD ، Steam

ترجع الزيادة في عدد الألعاب التي تم إصدارها على منصة Steam إلى الزيادة السريعة في التقدم التكنولوجي. يواجه المطورون ، وخاصة مطوري الألعاب المستقلة ، صعوبة في تحديد وقت الإصدار لأنه يتعين عليهم التنافس مع الألعاب الأخرى التي تم إصدارها في وقت واحد . و الهدف من هذا البحث العلمى هو التنبؤ بعدد إصدارات الألعاب على منصة Steam باستخدام انحدار KNN. البيانات المستخدمة حتى عام ١٨٢٦ هي بيانات ثانوية من ٢٠١٨ إلى ٢٠٢٢ تم تنزيلها من صفحة Kaggle.com وتجميعها في بيانات السلاسل الزمنية. الطرق المنفذة هي إعداد البيانات وتحويل البيانات ومعالجة البيانات وتقييم الأساليب المستخدمة. تطبيق البحث الأمثل عن المعلومات ، أي استخدام عدد متغيرات الإدخال من ١ إلى ١٤ متغيرا ، وانحدار KNN من ك=٣ إلى ك=٢٠ ينتج قيم MAD في سبعة أنواع من البيانات على النحو التالي: تحتوي بيانات جميع الألعاب على قيمة MAD تبلغ ٥.٦٤ عند ك=١٩ وثمانية متغيرات إدخال ؛ بيانات العمل ٢.٣٤ ؛ بيانات المغامرة ٢.٤٦ ؛ البيانات غير الرسمية ٢.٢٣ ؛ بيانات المحاكاة ١.٦٠ وبيانات آر بي جي ١.٦٤ حيث تحتوي جميع البيانات الخمسة على نفس المعلومات عند ك=٢٠ و ١٤ متغير إدخال ؛ وبيانات الإستراتيجية ١.٦٠ عند ك=١٩ و ١٤ متغيرات الإدخال. استنادا إلى أصغر قيمة تم الحصول عليها من MAD ، يمكن استنتاج أنه يمكن استخدام هذه الدراسة للتنبؤ بعدد الألعاب التي تم إصدارها على منصة Steam.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Peningkatan kemajuan teknologi yang pesat, juga memberikan pengaruh terhadap perkembangan *game* secara signifikan. Penjualan *game* yang awal mulanya berbentuk fisik kemudian berubah menjadi digital sehingga memberikan kemudahan pada banyak platform yang memberikan layanan ini. Salah satunya platform Steam yang menjadi incaran para penikmat *game* dalam memantau rilisan terbaru (Juwiantho, 2022). Steam merupakan *web e-commerce* dari Valve Corporation yang menyediakan ribuan *game* dari berbagai *publisher* secara digital dan diakses secara *online* melalui internet. Bahkan pada penelitian Grewal tahun 2022 menyatakan bahwa Steam merupakan platform terbesar dalam pendistribusian *game* secara daring di dunia.

Dengan adanya kemudahan ini, jumlah *game* yang dirilis dari tahun ke tahun terus mengalami peningkatan. Munculnya studi independen dan aksesibilitas yang besar terhadap perkembangan *indie games* memudahkan platform seperti Steam untuk meningkatkan jumlah *game* yang rilis setiap harinya, tercermin pada laman Steamspy yang memuat berbagai informasi mengenai hal tersebut (Pérez-Marcos et al., 2019). Banyaknya jumlah *game* yang rilis, menyebabkan para pengembang *game* kesulitan dalam menentukan waktu yang tepat untuk meluncurkan sebuah *game*, khususnya pengembang *indie game* karena harus bersaing dengan beberapa *game* dari *publisher* ternama yang rilis bersamaan di Steam. Sehingga, untuk mengantisipasi terjadinya penumpukan saat perilisan *game* di waktu yang sama

maka dibutuhkan sebuah peramalan untuk mengetahui jumlah *game* yang akan rilis pada waktu tersebut.

Peramalan adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil (Yuwantoro et al., 2019). Dalam matematika ada beberapa metode yang digunakan dalam meramal atau memprediksi sesuatu, salah satunya adalah algoritma *K Nearest Neighbour Regression*, yang akan digunakan pada penelitian ini. Pemilihan metode ini berdasarkan penelitian Yuwantoro pada tahun 2019 dalam memprediksi harga beras, hasil evaluasi *KNN Regression* jika dibandingkan dengan metode lain, menghasilkan nilai *root mean square error* atau RMSE yang lebih kecil jika ditemukan *k* yang paling tepat dan optimal. Algoritma *K-Nearest Neighbour Regression* menggunakan klasifikasi yang didasarkan pada ketetanggaan terdekat sebagai nilai prediksi dari sampel uji baru, jaraknya dihitung menggunakan rumus perhitungan jarak, salah satunya adalah *Euclidean Distance*.

Terdapat beberapa penelitian terdahulu yang mirip dengan penelitian ini. Penelitian yang dilakukan oleh (Hamdi et al., 2019) berkaitan tentang prediksi barang keluar, metode ini menghasilkan nilai evaluasi MAD yang cukup baik di bawah angka 1, akan tetapi terdapat nilai 0 dikarenakan kurangnya data yang diperoleh. Selanjutnya, penelitian oleh (Seruni et al., 2020), pada prediksi jumlah penduduk kota malang, beliau melakukan pencarian pada jumlah variabel bebas dan *k* - nya dengan batasan sebanyak 1 - 10, sehingga didapatkan nilai evaluasi MAPE yang terkecil, akan tetapi ketika diujikan dengan data baru terdapat kenaikan

nilai tersebut walaupun tidak signifikan. Kemudian, (Drajana, 2018) pada penelitian prediksi jumlah produksi minyak kelapa tahun 2018, beliau melakukan penambahan *feature selection* dengan menggunakan *backward elimination* untuk mencari nilai parameter atau variabel input dengan mencari nilai RMSE terkecil dengan batasan 1 – 5. Adapun pada penelitian (Nurfauzan & Fatimah, 2022) berkaitan dengan prediksi harga saham, penelitian tersebut menghasilkan akurasi yang cukup sempurna yaitu sebesar 98% dengan mencari nilai *k* terbaik menggunakan metode *elbow*. Dari beberapa penelitian sebelumnya, dengan kesamaan pada data *time series*, diharapkan algoritma KNN *Regression* mampu memberikan hasil yang baik pada penelitian ini.

Dengan judul “Prediksi jumlah rilis *game* pada platform Steam dengan menggunakan algoritma KNN *Regression*” yang berkaitan erat dengan waktu yang ditentukan, Al Quran juga menjelaskan tentang waktu yang tepat untuk melaksanakan haji pada surat Al-Baqarah ayat 189 yang berbunyi (Shihab, 2021):

يَسْأَلُونَكَ عَنِ الْأَهْلِ ۗ قُلْ هِيَ مَوَاقِيتُ لِلنَّاسِ وَالْحَجِّ ۗ وَلَيْسَ الْبِرُّ بِأَنْ تَأْتُوا الْبُيُوتَ مِنْ ظُهُورِهَا وَلَكِنَّ
الْبِرَّ مَنِ اتَّقَىٰ وَأَتُوا الْبُيُوتَ مِنْ أَبْوَابِهَا ۗ وَاتَّقُوا اللَّهَ لَعَلَّكُمْ تُفْلِحُونَ

Artinya:

“Mereka bertanya kepadamu (Nabi Muhammad) tentang bulan sabit. Katakanlah, “Itu adalah (penunjuk) waktu bagi manusia dan (ibadah) haji.” Bukanlah suatu kebajikan memasuki rumah dari belakangnya, tetapi kebajikan itu adalah (kebajikan) orang yang bertakwa. Masukilah rumah-rumah dari pintu-pintunya, dan bertakwalah kepada Allah agar kamu beruntung”(QS. Al-Baqarah;/4:58).

Dari ayat di atas terdapat bulan sabit yang diartikan sebagai pergantian bulan, dalam penentuan waktu ibadah seperti puasa Ramadhan dan haji. Sehingga secara tidak langsung, ayat tersebut berkaitan dengan penelitian ini karena dengan mengetahui jumlah *game* yang akan rilis di Steam pada waktu kedepannya, akan mempermudah dalam mempersiapkan diri.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan di atas, rumusan masalah dalam penelitian ini adalah bagaimana hasil evaluasi berdasarkan prediksi jumlah rilis *game* dengan Algoritma *KNN Regression* pada platform Steam.

1.3 Tujuan Penelitian

Merujuk pada rumusan masalah di atas, maka tujuan dan penulis skripsi ini adalah untuk mengetahui hasil evaluasi berdasarkan prediksi jumlah rilis *game* Algoritma *KNN Regression* pada platform Steam.

1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian maka diharapkan beberapa manfaat sebagai berikut:

1. Bagi Penulis

Penelitian ini diharapkan dapat memberikan wawasan mengenai algoritma *KNN Regression* untuk model jumlah rilis *game* pada platform Steam dan dapat berguna untuk penelitian selanjutnya.

2. Bagi Mahasiswa

Penelitian ini diharapkan dapat menambah pengetahuan berkaitan komputasi matematika khususnya tentang keterkaitan dengan algoritma *KNN Regression* dalam membuat model prediksi dari data yang digunakan pada penelitian ini.

3. Bagi Instansi

Hasil penelitian ini dapat digunakan sebagai tambahan bahan literatur atau rujukan dan bahan pengembangan ilmu komputasi dalam matematika. Khususnya yang berhubungan dengan algoritma *KNN Regression*.

4. Bagi Pembaca

Penelitian ini diharapkan dapat bermanfaat untuk masyarakat secara umum, dan secara khusus untuk para developer maupun jurnalis agar dapat mengetahui jumlah rilis *game* pada platform Steam sehingga dapat mempersiapkan waktu yang tepat untuk merilis *game* maupun konten.

1.5 Batasan Masalah

Batasan masalah dimaksudkan untuk memenuhi tujuan penelitian dan tidak terjadi perluasan masalah antara lain

1. Pencarian model terbaik berdasarkan nilai MAD terendah pada penggunaan jumlah variabel input sebanyak 1 hingga 14 variabel, nilai k (*K-Nearest Neighbor*) antara 3 sampai 20, dan $k - Fold$ dengan 10 *Fold*.
2. Data yang digunakan pada penelitian ini merupakan data *game* Steam dari tahun 2018 – 2022.
3. Variabel terdiri dari jumlah rilis harian yang disusun secara terurut berdasarkan data *time series*.

1.6 Definisi Istilah

Metode : Prosedur dalam menyelesaikan masalah-masalah yang diteliti atau dihadapi.

- MAPE : Ukuran kesalahan *relative*
- RMSE : Pengukuran berdasarkan nilai yang berbeda dari observasi pada model regresi, semakin kecil nilainya, semakin bagus (Nurfauzan & Fatimah, 2022).
- Algoritma : Algoritma adalah suatu upaya dengan urutan operasi yang disusun secara logis dan sistematis untuk menyelesaikan suatu masalah untuk menghasilkan suatu output tertentu.
- Model : Hasil dari algoritma yang diterapkan pada kumpulan data

BAB II

KAJIAN TEORI

2.1 *Data Time Series*

Data merupakan bentuk jamak dari datum, yang berarti suatu hal yang memiliki keterangan–keterangan mengenai sesuatu yang diketahui, dapat berupa anggapan atau fakta dalam bentuk angka maupun simbol-simbol yang lain (Hasan, 2006). Secara garis besar, data adalah penulisan berulang mengenai karakteristik suatu objek dengan masing-masing objek memiliki identitas tersendiri yang didefinisikan ke dalam variabel (Saefudin et al., 2009). Terdapat dua jenis data yaitu data primer dan data sekunder, yang di mana pada penelitian ini menggunakan data sekunder. Data sekunder adalah data yang diperoleh dari sumber-sumber yang telah tersedia, di mana data tersebut telah dikumpulkan oleh peneliti sebelumnya. Data ini biasanya ditemukan di perpustakaan, laporan penelitian terdahulu, *database* daring dan lain-lain (Hasan, 2006).

Data juga dibagi menjadi beberapa bentuk, salah satunya adalah data *time series* atau data berkala. Data berkala adalah data yang dikumpulkan dari waktu ke waktu yang menunjukkan gambaran atau perkembangan dari suatu objek data (Hasan, 2006). Data tersebut dikumpulkan beberapa kali pada interval waktu yang relatif sama, memiliki kesamaan pada *instrument* maupun objeknya. Model *time series* pada penelitian eksperimen akan menghasilkan data *time series* pula (Sugiyono, 2018).

Data *time series* dapat dikaitkan sebagai analisis deret waktu dan dibagi menjadi dua bagian yaitu *one-step-ahead* dan *multi-step-ahead*, contohnya pada

peramalan dengan analisis deret waktu sederhana yang berhubungan dengan *one-step-ahead* atau satu langkah maju diberikan formula sebagai berikut (Prado & West, 2010):

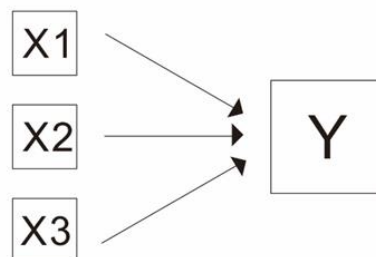
$$x_{t+1} = x_t + x_{t-1} + x_{t-2} + \dots + x_1 \quad (2.1)$$

Di mana x_{t+1} adalah waktu esok, x_t merupakan waktu saat ini, x_{t-1} adalah waktu satu hari yang lalu, x_{t-2} adalah waktu dua hari yang lalu dan seterusnya hingga beberapa hari lalu yang ditentukan.

2.2 Regresi

Regresi merupakan salah satu model dari statistika yang di mana melihat perubahan nilai suatu variabel terhadap variabel yang lain. Tujuan dari regresi adalah untuk membuat prakiraan atau prediksi nilai suatu hal ketika diketahui nilai yang mempengaruhinya. Model regresi yang memiliki hanya satu variabel terikat disebut dengan regresi sederhana (*Simple Regression*), sedangkan model regresi yang memiliki lebih dari satu variabel terikat disebut regresi berganda (*Multiple Regression*) (Subagyo, 2010). Pada dasarnya, analisis regresi merupakan analisis yang memprakirakan kemungkinan perubahan pada variabel terikat y yang berkaitan dengan nilai tertentu dari variabel bebas X (Scheffler, 1987). Penyajian dalam bentuk gambar, model regresi akan terlihat seperti berikut.

Analisis Regresi
Spesifikasi Model



Gambar 2.1 Model Regresi

Model di atas menunjukkan bahwasanya variabel terikat Y dipengaruhi oleh variabel bebas x_1, x_2 , dan x_3 secara bersama-sama maupun secara parsial. Penjelasan mengenai variabel dalam penelitian ini dibagi menjadi dua jenis (Fauzi, 2009).

1. Variabel Tergantung

Variabel tergantung atau variabel target, variabel ini sering disebut variabel output, kriteria atau *dependent*, merupakan variabel penelitian yang bertujuan untuk mengukur seberapa besar efek atau pengaruh yang diberikan variabel lain. Pada penelitian ini, hanya 1 variabel target yang akan digunakan, berupa waktu setelahnya dari data *time series*.

2. Variabel Bebas

Variabel bebas atau variabel input, variabel ini sering disebut juga sebagai variabel stimulus, prediktor, atau *independent*, merupakan variabel penelitian yang variasinya memberikan pengaruh terhadap variabel yang lain. Penelitian ini menggunakan 1 hingga 14 variabel bebas berdasarkan waktu harian, yang disusun secara terurut dalam bentuk data *time series*.

2.3 *Machine Learning*

Machine learning atau pembelajaran mesin merupakan salah satu bagian dalam kecerdasan buatan yang mempelajari suatu sistem dengan memasukan algoritma dan data – data agar mampu melakukan tugasnya sendiri tanpa dibantu oleh penggunanya. Dalam hal ini, memungkinkan programmer dapat belajar (Yuwantoro et al., 2019). Salah satu algoritma dalam *machine learning* adalah *K-Nearest Neighbor* yang dapat digunakan pada klasifikasi maupun regresi.

2.3.1 Algoritma *K-Nearest Neighbor Regression*

Secara sederhana, algoritma *K-Nearest Neighbor* merupakan salah satu algoritma dalam *data mining* yang digunakan pada klasifikasi maupun regresi berdasarkan kedekatan jarak antar satu data dengan data yang lain (Anggie, 2018). Adapun Algoritma *K-Nearest Neighbor Regression* melakukan prediksi berdasarkan nilai k tetangga atau data terdekat dengan nilai nilai data yang lainnya. Oleh karena itu, untuk mengukur jarak antara satu titik data ke data yang lainnya dibutuhkan metode matematika yang cocok, salah satunya adalah *Euclidean distance* (Tanuwijaya & Hansun, 2019).

Untuk lebih jelasnya, di bawah ini merupakan langkah – langkah dalam menerapkan Algoritma *K Nearest Neighbor* (Mulyati et al., 2020):

1. Menentukan nilai k
2. Menghitung jarak setiap data *testing* dengan semua data *training*
3. Mengurutkan setiap nilai dari jarak antar data
4. Mengelompokkan nilai data *training* yang terkecil sebanyak k data
5. Menjumlahkan semua nilai target dari data *training* yang dikelompokkan sebelumnya, kemudian membaginya sebesar nilai k dengan rumus berikut:

$$\hat{f}(x_{test(i)}) = \frac{\sum_{i=1}^k w_i f(x_{train(i)})}{\sum_{i=1}^k w_i} \quad (2.2)$$

Keterangan:

$\hat{f}(x_{test(i)})$: Nilai prediksi

w : Beban, bernilai 1 jika tidak berkaitan

$f(x_{train(i)})$: Nilai target atau aktual dari data training

k : k dari *KNN Regression* yang telah ditentukan

i : Iterasi

2.4 Perhitungan Jarak

Jarak merupakan aspek yang penting dalam penelitian berbasis klasifikasi maupun regresi. Konsep jarak sendiri, memberikan banyak pengembangan terhadap berbagai metode dalam literatur *machine learning* atau *data mining* (Santosa, 2007). Ukuran jarak juga harus memenuhi syarat-syarat berikut:

1. $d(x_{test(i)}, x_{train(i)}) \geq 0$ (*non-negative*)

Tidak ada jarak yang memiliki nilai negatif

2. $d(x_{test(i)}, x_{train(i)}) = 0$, jika dan hanya jika $x_{test(i)} = x_{train(i)}$

Jarak antara data *training* ke- i dengan data *testing* ke- i bernilai 0

3. $d(x_{test(i)}, x_{train(i)}) = d(x_{test(i)}, x_{train(i)})$

Jarak dari data *testing* ke- i ke data *training* ke- i bersifat simetri

4. $d(x_{test(i)}, x_{train(i)}) \leq d(x_{test(i)}, x_{train(i)}) + d(x_{train(i)}, x_{test(i)})$

Rumus ketidaksamaan segitiga

2.4.1 Euclidean Distance

Euclidean Distance merupakan metode paling populer yang sangat cocok untuk digunakan pada prediksi secara numerik (Tanuwijaya & Hansun, 2019). Metode ini mengukur secara garis lurus ataupun langsung kedekatan antara dua buah objek (data) dan dihitung dengan persamaan berikut.

$$d_{Euclidean}(x_{train(i)}, x_{test(i)}) = \sqrt{\sum_{i=1}^p (x_{train(i)} - x_{test(i)})^2} \quad (2.3)$$

Keterangan:

$d_{Euclidean}$: Jarak *Euclidean*

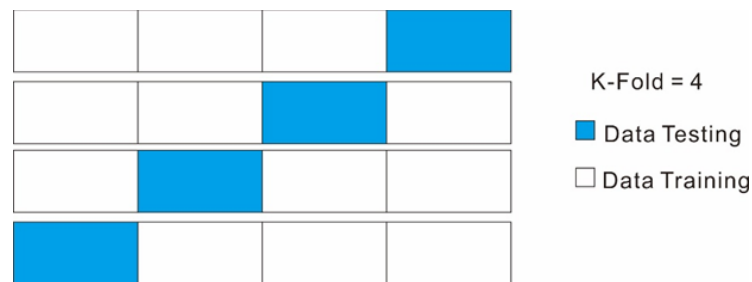
$x_{test(i)}$: Data *testing* ke- i

$x_{train(i)}$: Data *training* ke- i

p : Panjang variabel input atau panjang data

2.5 *K-Fold Cross Validation*

Cross Validation adalah salah satu metode yang digunakan untuk memperkirakan akurasi kinerja model prediksi dengan membagi data *training* dan *testing* secara acak. Metode ini paling banyak digunakan ketika target adalah prediksi (Andini et al., 2022). Adapun *K-Fold Cross Validation* merupakan salah satu pengujian data yang melakukan perulangan dengan membagi data sejumlah n -*fold* atau n -lipatan tertentu. Pada prosesnya, data akan dibagi kedalam n buah partisi menjadi beberapa data yang ukurannya sama sebanyak n_1, n_2, n_3 dan seterusnya, kemudian melakukan proses *training* dan *testing* sebanyak n kali pada partisi iterasi ke-1 akan menjadi data *testing*, dan iterasi yang lainnya akan menjadi data *training*, hingga data *testing* menjadi iterasi yang terakhir (Juwiantho, 2022).



Gambar 2.2 *K-Fold Cross Validation*

Pada **Gambar 2.2** menunjukkan $K - Fold = 4$ yang berarti jumlah data akan dibagi menjadi 4 lipatan dengan $K = n$ dimana $n = 1,2,3,4$. Pada pembagiannya, 1 lipatan akan menjadi data *testing* dan 3 lipatan akan menjadi data

training yang kemudian akan dilakukan perulangan sebanyak 4 kali dengan data *testing* berada pada lipatan ke-1 hingga berada pada lipatan ke-4.

2.6 *Lost Function*

Modifikasi Dalam regresi ada beberapa perhitungan *error* yang sering digunakan untuk menilai performa suatu model. Jika nilai prediksi \hat{y} untuk data ke- i dan nilai y sebagai data aktual dari data ke- i , maka beberapa ukuran *error* yang sering dipakai terdapat pada fungsi *lost function* (Santosa, 2007). *Lost function* merupakan sebuah fungsi untuk mengukur seberapa baik model saat ini berdasarkan hasil prediksi dengan hasil yang diharapkan pada algoritma yang digunakan.

2.6.1 MAD

MAD atau *Mean Absolute Deviation* merupakan metode untuk menghitung akurasi yang ditentukan seberapa baiknya berdasarkan hasil terkecil dari keluaran model (Hamdi et al., 2019). MAD atau rata-rata deviasi mutlak menghitung rata-rata deviasi dengan mengabaikan tanda positif atau negatif dan menggunakan nilai mutlak untuk menghitungnya (Ansori, 2020).

$$MAD = \frac{\sum_{i=1}^{X_{test}} |f(x_{test(i)}) - \hat{f}(x_{test(i)})|}{X_{test}} \quad (2.4)$$

Keterangan:

$f(x_{test(i)})$: Nilai target dari data *testing* ke- i

$\hat{f}(x_{test(i)})$: Nilai prediksi dari data *testing* ke- i

X_{test} : Jumlah data *testing*

2.7 Kajian Integrasi Topik dengan Islam

Dalam memprediksi jumlah rilis *game*, hal ini tidak jauh kaitannya dengan waktu rilis dan jumlah suatu objek yang diteliti. Sebagaimana potongan ayat dalam firman Allah SWT dalam surat Al-A'raf ayat 86 yang berbunyi (Shihab, 2021):

وَأذْكُرُوا إِذْ كُنْتُمْ قَلِيلًا فَكَثَّرَكُمْ

“Dan ingatlah di waktu dahulunya kamu berjumlah sedikit, lalu Allah memperbanyak jumlah kamu”

Berdasarkan firman Allah SWT dalam surat Al-A'raf ayat 86, menjelaskan bahwa sesungguhnya Allah mengetahui jumlah manusia mengalami kenaikan seiring waktu berlalu. Dalam tafsir Quraish Shihab mengatakan, manusia harus mengingat waktu yang sebelumnya atau lampau bahwasanya mereka berjumlah sedikit, kemudian Allah jadikan jumlah manusia semakin banyak di tahun - tahun berikutnya berdasarkan pengaruh sebab-akibat.

2.8 Kajian Integrasi Topik dengan Teori Pendukung

Hasil prediksi pada penelitian ini, dapat dilihat dari seberapa besar nilai MAD yang diperoleh. Untuk mendapatkan nilai MAD terendah dibutuhkan pencarian parameter optimal berdasarkan penggunaan jumlah variabel input dan k dari KNN *Regression*. Perbedaan data *testing* dan data *training* juga dapat mempengaruhi nilai MAD, pada penelitian terdahulu, penelitian (Drajana, 2018) menggunakan metode *K-Fold Cross Validation* pada $k = 10$ memberikan hasil RMSE yang rendah dalam memprediksi jumlah produksi *coconut oil*. Kemudian, pada penelitian (Zulfallah, 2022) yang berkaitan dengan kelulusan mahasiswa, beliau menerapkan metode *K-Fold Cross Validation* pada $k = 10$ dalam mencari

parameter yang optimal, sehingga berdasarkan hasil tersebut, penelitian ini menggunakan $k - Fold = 10$ pada pembagian datanya.

Adapun pada penggunaan jumlah variabel input, penelitian (Seruni et al., 2020) menggunakan pencarian dari 1 hingga 10 variabel input dalam bentuk bulanan untuk memprediksi jumlah pertumbuhan penduduk, hasil dengan nilai MAPE terbaik berada pada penggunaan 6 variabel input yang di mana terjadi kenaikan nilai MAPE untuk jumlah variabel input kurang dari 6 dan lebih dari 6. Jumlah variabel input yang sedikit belum menggambarkan pola dalam memprediksi pertumbuhan penduduk, sedangkan jumlah variabel lebih dari 6, tidak lagi menggambarkan hubungan antar data. Pada penelitian ini menggunakan 1 sampai 14 variabel input dalam bentuk harian untuk mencari nilai MAD yang paling optimal, penggunaan dibatasi sebanyak 14 variabel input dikarenakan nilai MAD di antara jumlah tersebut tidak terlalu memiliki perbedaan yang signifikan.

Penelitian (Zulfallah, 2022) untuk mengukur ketepatan kelulusan mahasiswa melakukan pencarian k dari KNN *Regression* yang di mana akurasi terbaik berada pada k ke-31, akurasi dari k yang besar dipengaruhi oleh jumlah data yang besar pula yaitu sebanyak 27000 lebih data. Jumlah k di antara k ke-31 memiliki nilai akurasi yang saling berdekatan, adapun pada penelitian (Seruni et al., 2020) pencarian k terbaik berada pada k ke-5 dengan jumlah data sebanyak 75 data, jumlah k yang kecil juga dapat dipengaruhi jumlah data yang sedikit. Sehingga berdasarkan hal di atas, pada penelitian ini menggunakan k ke-3 hingga ke-20 pada total data sebanyak 1826 data.

2.9 Steam

Steam adalah salah satu platform paling populer dalam menyediakan *game* secara digital, dengan menawarkan berbagai layanan seperti pembelian dan pengunduhan *game* pc, perlindungan hak cipta digital, *matchmaking* dalam *game* online, *streaming* video, forum jejaring social, dan lain sebagainya (Li & Zhang, 2020). Dalam platform Steam, *game* dibagi menjadi genre yang sangat bervariasi, beberapa genre *game* akan dipakai pada penelitian ini berdasarkan popularitasnya dan besarnya jumlah *game* (Grewal et al., 2022). Berikut adalah definisi dari genre *game* yang dipilih (Qaffas, 2020).

1. Aksi

Game action dalam bahasa Indonesia memiliki arti permainan dengan genre aksi, *game* ini membutuhkan ketangkasan dan keputusan yang cepat serta akurat pada setiap adegan sehingga membutuhkan fokus dan perhatian lebih (Kapp et al., 2020).

2. Petualangan

Game adventure atau petualangan merupakan permainan dengan mengandalkan cerita yang menarik di mana setiap perjalanan berhubungan dengan *gameplay*-nya. Kata ‘petualangan’ disini berkaitan dengan misi yang terstruktur atau alur yang linear.

3. Kasual

Game casual atau kasual dapat diartikan sebagai permainan yang memiliki control yang mudah dan aturan yang tidak kompleks pada *gameplay*-nya. Secara sederhana merupakan *game* dengan instruksi yang singkat dan sangat

mudah dipahami oleh para calon pemain, kegagalan dalam memainkan *game* ini tidak memberikan dampak yang berat terhadap pemainnya (Kapp et al., 2020).

4. Simulasi

Game simulation atau simulasi merupakan permainan yang menggunakan fenomena dunia nyata sebagai objek interaksi secara virtual.

5. Strategi

Game strategy atau strategi memiliki fitur utama yaitu keputusan yang kompleks yang dipilih para pemain dalam sebuah *game* bernuansa simulasi atau fantasi politik, perekonomian, atau dunia militer biasanya dengan suasana perang.

6. RPG

RPG merupakan singkatan dari *Role-Playing Game* merupakan permainan dengan latar dunia fantasi atau imajinasi. Para pemain memiliki kebebasan untuk memilih bagaimana cara mengeksplorasi dunia dalam *game* tersebut, pada suatu kondisi ketika menelusuri alur atau cerita maupun misi, kemungkinan mereka akan kembali ke tempat yang telah didatangi atau dieksplor. Ketersediaan *game* dalam menyediakan dunia untuk dieksplorasi, biasanya sangat luas sekali.

BAB III

METODE PENELITIAN

3.1 Pendekatan Penelitian

Penelitian dalam masalah ini menggunakan jenis penelitian eksperimen kuantitatif, karena metode eksperimen adalah metode penelitian kuantitatif yang berguna untuk melihat pengaruh variabel bebas terhadap variabel terikat pada kondisi yang ditentukan. Kondisi tersebut mengacu pada tidak adanya pengaruh yang diberikan selain variabel bebas. Dalam metode ini, variabel bebas dan terikat disusun berdasarkan deret waktu, dimana variabel bebasnya terdiri dari 1 hingga 14 hari, dan 1 variabel terikat yang merupakan hari setelahnya. Tujuan metode ini untuk memahami fenomena secara umum dengan menggambarkan fakta, mengembangkan, melakukan pembuktian dan menemukan pengetahuan (Sugiyono, 2018).

Model kuantitatif merupakan model keputusan yang menggunakan angka. Pada penelitian ini, angka mempunyai peranan yang sangat penting dalam pembuatan, penggunaan, dan pemecahan model kuantitatif. Setiap variabel bebas, variabel terikat, maupun output dari model pada penelitian ini, merupakan angka. Pada dasarnya penggunaan model kuantitatif untuk memecahkan masalah maupun keputusan keputusan dalam suatu model, hasil akhir yang keluar juga berupa angka (Muslich, 2009). Secara garis besar, metode kuantitatif dapat diartikan sebagai metode yang data penelitiannya berupa angka dan dihitung menggunakan ilmu statistika (Imron, 2019).

3.2 Data dan Sumber Data

Pada bab sebelumnya, sudah dijelaskan apa itu data dan data yang diambil berasal dari sumber data. Yang dimaksud sumber data disini adalah suatu subyek dari mana data di peroleh (Ansori, 2020). Untuk data yang dipakai pada penelitian ini merupakan data *game* yang berisi informasi berkaitan nama, *publisher*, perilsan dan lain-lain dalam platform Steam yang diambil dari situs kaggle.com di mana situs tersebut berisi berbagai macam data. Data tersebut akan dipilah Kembali berdasarkan informasi perilsan *game*.

Tabel 3.1 Sample Data Steam dari Kaggle.com

Link	Nama	Tahun Rilis	Genre
66403	<i>Space Empires IV Deluxe</i>	07/02/2006	<i>['Strategy', '4X', 'Sci-fi', 'Space', 'Turn-Based Strategy', 'Turn-Based', 'Singleplayer']</i>

Data **Tabel 3.1** merupakan contoh sebagian data dari data mentah yang belum diolah menjadi data *time series*, data yang lebih lengkap dapat dilihat pada halaman lampiran, berikut merupakan jenis data yang akan diproses pada penelitian ini.

Tabel 3.2 Data *Time Series* pada $p = 7$

p1	p2	p3	p4	p5	p6	p7	target
13	10	13	18	23	6	0	10
10	13	18	23	6	0	10	16

13	18	23	6	0	10	16	17
18	23	6	0	10	16	17	17

3.3 Teknik Analisis Data

Analisis data pada dasarnya berarti menentukan atau memperkirakan seberapa besar pengaruh kejadian terhadap kejadian lainnya secara kuantitatif dan memperkirakan atau meramalkan kejadian tersebut, kejadian disini dapat diartikan sebagai perubahan nilai variabel atau objek dalam data (Hasan, 2006). Secara umum dalam penelitian ini, terdapat beberapa langkah yang digunakan agar dapat melakukan proses dan pengujian terhadap model yang ditentukan.

1. Persiapan Data

Dataset yang telah dikumpulkan, akan diambil dan dipilah informasi yang sesuai dengan model yang dibuat, dalam hal ini penulis mengambil informasi atau bisa disebut *reducing data* berfokus pada waktu rilis, dan genre *game* yang diambil dari tahun 2018 hingga 2022 dengan total data sebanyak 1826.

2. Transformasi Data

Setelah pembentukan *dataset*, maka akan dilakukan perubahan *dataset* menjadi data *time series* dengan waktu harian, agar data tersebut dapat diolah menjadi model prediksi pada penelitian ini. *Dataset* kemudian akan dibagi menjadi data input dan data target, di mana data inputnya atau variabel bebas akan dibatasi sebanyak 14 data berdasarkan jumlah rilis perhari secara terurut dan hanya 1 data target atau variabel terikat dengan data 1 hari setelahnya.

3. Pemrosesan Data

Pada pembentukan model dengan menggunakan algoritma *KNN Regression*, diberikan beberapa langkah berikut:

a. Pemilihan Data

Langkah pertama, data dipilih berdasarkan jumlah variabel input antara 1 hingga 14 variabel.

b. Pembagian Data

Pada tahap ini, data dibagi menjadi dua bagian yaitu data *testing* dan *training* dengan metode *K-Fold Cross Validation* sebanyak 10 *fold*.

c. Perhitungan Jarak

Setelah tahap pembagian data, tahap selanjutnya yaitu menghitung setiap jarak dari masing-masing data *testing* ke setiap data *training* pada 10 *fold* menggunakan *Euclidean Distance*. Data kemudian diurutkan berdasarkan jarak terdekat atau nilai terkecil yang telah didapatkan.

d. Penerapan Algoritma *K-Nearest Neighbor Regression*

Setelah jarak dari setiap data *testing* ke seluruh data *training* diurutkan, selanjutnya menentukan parameter k dari *K-Nearest Neighbor Regression* antara $k = 3$ hingga $k = 20$. Kemudian mengambil nilai target dari data *training* berdasarkan jarak terdekat terhadap masing-masing data *testing* sebanyak parameter yang telah ditentukan. Selanjutnya menjumlahkan nilai target dari data *training* yang telah diambil, kemudian membaginya dengan sejumlah parameter k . Hasil pembagian inilah yang akan menjadi nilai prediksi dari data *testing*.

e. Menghitung Nilai MAD

Setelah mendapatkan nilai prediksi, nilai tersebut akan dibandingkan dengan nilai target dari data *testing* dengan menggunakan perhitungan nilai *error* yaitu MAD atau *Mean Absolute Deviation*. Semakin kecil nilai MAD, maka semakin akurat nilai prediksinya.

f. Pencarian Parameter

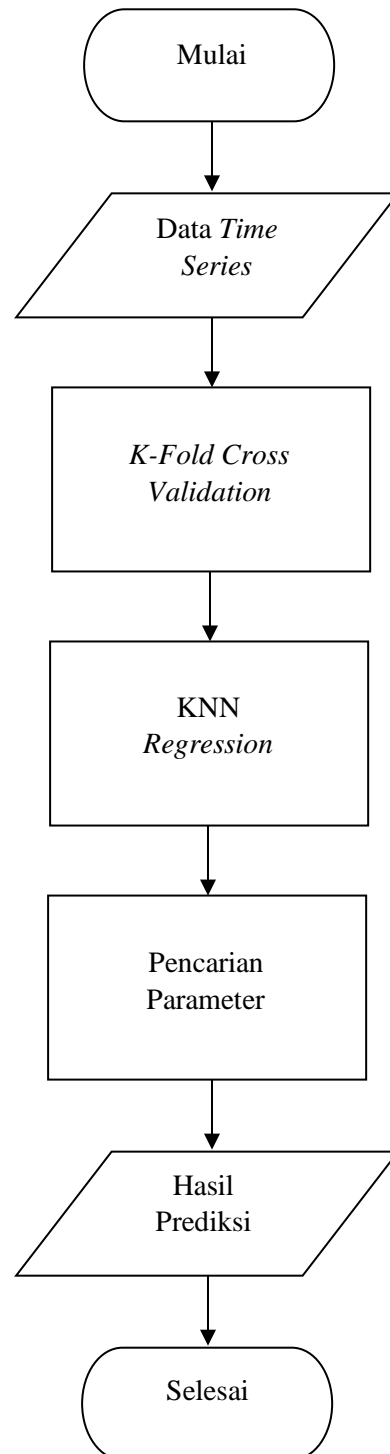
Untuk menemukan performa terbaik dari model, maka diperlukan pencarian pada parameter-parameter yang mempengaruhinya. Pada model Algoritma KNN *Regression*, pencarian parameter didasarkan pada jumlah variabel input antara 1 hingga 14 variabel dan parameter $k = 3$ hingga $k = 20$ dari KNN *Regression*, model dengan parameter yang memberikan nilai MAD terkecil akan diujikan terhadap 14 data baru.

g. Pengujian Model

Model dengan performa terbaik, diujikan terhadap data baru untuk melihat bagaimana perubahan yang dihasilkan dari model tersebut melalui selisih antara hasil prediksi dan nilai target, yang kemudian akan ditentukan nilai MAD dari data baru tersebut.

3.4 *Flowchart* Perhitungan Prediksi

Gambar 3.1 *Flowchart* Perhitungan Prediksi



BAB IV

PEMBAHASAN

4.1 Persiapan Data

Data Steam *game* yang telah dikumpulkan sebelumnya akan diolah terlebih dahulu untuk dapat diproses pada tahap berikutnya. Pengolahan data dilakukan dengan cara mengambil sebagian informasi yang diperlukan pada pembuatan model prediksi. Informasi ini berupa tanggal perilsan *game* yang dimulai pada tahun 2018 hingga 2022, kemudian data tersebut akan dibagi menjadi 6 genre yang dipilih berdasarkan tingkat popularitasnya. Pembagian ini dilakukan untuk melihat seberapa banyak *game* yang dirilis pada masing – masing genre yang telah ditentukan yaitu aksi, petualangan, kasual, simulasi, strategi dan RPG. Data dikelompokkan berdasarkan jumlah rilis perhari sehingga total data keseluruhan berjumlah sebanyak 1826 data. Berikut akan ditunjukkan contoh sebagian data berdasarkan tanggal rilis dan jumlah pada **Tabel 4.1** dan data lebih lengkap pada halaman lampiran:

Tabel 4.1 Jumlah Data Semua *Game* yang Rilis Perhari

Tanggal Rilis	Jumlah
1 Januari 2018	13
2 Januari 2018	10
3 Januari 2018	13
4 Januari 2018	18

4.2 Transformasi Data

Pada tahap ini, data Steam *game* akan diubah menjadi data berbentuk *time series* atau deret waktu, dimana data tersebut akan dikelompokkan berdasarkan jumlah *game* yang rilis setiap harinya. Data akan disusun dalam bentuk *time series* yang terdiri dari dua variabel, yaitu variabel input dari 1 hingga 14 hari secara terurut dan variabel targetnya merupakan jumlah rilis *game* pada 1 hari selanjutnya. Variabel input merupakan variabel bebas yang diberi label p dan variabel target sebagai variabel terikat akan diberi label target. Secara detail, label p1 berarti jumlah *game* yang rilis pada hari pertama, p2 merupakan jumlah *game* yang rilis pada hari setelah p1, atau hari kedua dan ini berlaku hingga label target.

Tabel 4.2 Contoh Data Semua *Game* dalam bentuk *time series* dengan 7 Variabel Input

p1	p2	p3	p4	p5	p6	p7	target
13	10	13	18	23	6	0	10
10	13	18	23	6	0	10	16
13	18	23	6	0	10	16	17
18	23	6	0	10	16	17	17
23	6	0	10	16	17	17	24
6	0	10	16	17	17	24	4

Pada **Tabel 4.2**, data nomor 2 dengan label p1 di isi kembali dengan nilai yang sama pada label p2 nomor 1, label p2 diisi dengan label p3 nomor 1, dan seterusnya hingga label target. Label target kemudian akan dibandingkan dengan hasil prediksi, yang nilai prediksinya didasarkan pada data dengan urutan 1 hingga 14 hari sebelumnya. Berikut akan ditunjukkan dalam tabel, berapa jumlah data yang akan diproses untuk setiap p nya.

Tabel 4.3 Jumlah Data berdasarkan Jumlah Variabel input

Jumlah Penggunaan Variabel Input	Jumlah Data
1 variabel input	1825
2 variabel input	1824
3 variabel input	1823
4 variabel input	1822
5 variabel input	1821
6 variabel input	1820
7 variabel input	1819
8 variabel input	1818
9 variabel input	1817
10 variabel input	1816
11 variabel input	1815
12 variabel input	1814
13 variabel input	1813
14 variabel input	1812

Dapat dilihat pada **Tabel 4.3**, jumlah data berkurang sebanyak 1 data disetiap kenaikan jumlah variabel input. Data dengan 1 variabel input yaitu p1, memiliki jumlah data sebanyak 1825, sedangkan pada penggunaan 2 variabel input, p1 dan p2, jumlah data mengalami pengurangan sebesar 1 buah data. Pengurangan ini terjadi karena banyaknya penggunaan variabel input, maka susunan dari data *time series* semakin panjang. Hal ini juga menyebabkan nilai pada masing-masing variabel target disetiap data antara p1 hingga p14 berbeda.

4.3 Pemrosesan Data

Data yang sudah di transformasi, kemudian akan diproses dalam pembuatan model prediksi menggunakan algoritma *KNN Regression*.

1. Pemilihan Data

Memilih data antara 1 variabel input (p_1), atau 2 variabel input (p_1 dan p_2), atau 3 variabel input (p_1 , p_2 , dan p_3), hingga 14 variabel input (p_1 , p_2, \dots, p_{14}) sebagai proses awal. Disini akan dipilih data dengan 7 variabel input terlebih dahulu, kemudian akan diproses pada tahap selanjutnya. Pada dasarnya, data dengan 1 variabel input hingga 14 variabel input, semuanya akan diujikan dan dipilih berdasarkan nilai MAD atau *mean absolute deviance* terkecil.

2. Pembagian Data

Data sebanyak 1819 tersebut kemudian akan dibentuk menjadi data *training* dan data *testing* menggunakan metode *K-Fold Cross Validation* sebanyak 10 lipatan atau 10 *fold*, yang artinya data model akan dibagi sebesar 10 bagian, 1 bagian sebagai data *testing* dan 9 bagian sebagai data *training*.

Data uji / *testing*:

$$\begin{aligned} X_{test} &= \frac{X}{N} & (4.1) \\ &= \frac{1819}{10} \\ &= 181.9 \end{aligned}$$

Hasil dari pembagian data harus berupa bilangan bulat positif, sehingga berdasarkan perhitungan pembagian data *testing* yang ditunjukkan pada rumus (4.1), hasil sebesar 181.9 akan diterapkan pada setiap partisi dengan jumlah 182 data *testing* untuk partisi ke-1 dan ke-9, sedangkan partisi ke-10

berjumlah 181 data *testing*. Hal ini juga berlaku untuk pembagian pada data *training*.

Data *training*:

$$\begin{aligned} X_{train} &= X - X_{test} & (4.2) \\ &= 1819 - 181.9 \\ &= 1637,1 \end{aligned}$$

Keterangan:

- X_{test} : Jumlah data *testing*
- X : Jumlah data
- N : Jumlah total partisi pada *K-Fold Cross Validation*

Fold pertama memiliki data *testing* dari data ke-1 hingga ke-182, dan data *training* dari data ke-183 hingga data ke-1819, kemudian *fold* kedua dengan data *testing* dari data ke-183 hingga ke-364, dan data *training*-nya adalah data ke-1 hingga data ke-182 dan data ke-365 hingga data ke-1819, begitu pun seterusnya hingga *fold* kesepuluh.

3. Perhitungan Jarak

Menghitung jarak data *testing* pada setiap *n-Fold* ke seluruh data *training* dengan menggunakan *Euclidean distance*. Berikut contoh perhitungan jarak secara manual menggunakan rumus (2.2) untuk data *testing* ke-1 pada *n-Fold* ke-1 terhadap 3 data *training* dari urutan pertama dan menggunakan program Python untuk menghitung jarak *Euclidean* terhadap keseluruhan data *training* yang ditunjukkan pada halaman lampiran:

$$\begin{aligned} d(x_{l1}, x_{u1}) &= \sqrt{(29 - 13)^2 + (13 - 10)^2 + (13 - 13)^2 + \dots + (0 - 0)^2} \\ &= 19,75 \end{aligned}$$

$$d(x_{12}, x_{u1}) = \sqrt{(13 - 13)^2 + (13 - 10)^2 + (23 - 13)^2 + \dots + (15 - 0)^2} \\ = 29,7$$

$$d(x_{13}, x_{u1}) = \sqrt{(13 - 13)^2 + (23 - 10)^2 + (33 - 13)^2 + \dots + (30 - 0)^2} \\ = 47,1$$

Mengurutkan setiap nilai jarak antara data *testing* ke-1 terhadap seluruh data *training* dari yang paling terkecil. Seperti pada **Tabel 4.4**, data ke 372 memiliki nilai jarak terkecil di urutan pertama dengan nilai targetnya adalah 5, kemudian data ke 358 di urutan kedua dengan nilai targetnya 11, dan seterusnya, hingga data yang berada pada urutan terakhir, dengan nilai jarak terbesar.

Tabel 4.4 Urutan Terkecil Jarak Data *Test* ke - 1

Data	Jarak	Target
372	11,83	5
358	11,96	11
400	12,65	17
736	13,00	21
750	14,07	20
722	15,03	10
547	16,40	4
721	16,64	7
442	17,00	17
729	17,41	20
526	18,17	24
323	18,22	18
540	18,73	12
546	18,87	1
1457	19,24	18
720	19,49	4
183	19,75	15

4. Penerapan Algoritma *K-Nearest Neighbor Regression*

Setelah mengurutkan nilai jarak dari data *testing* ke-1 terhadap data *training*.

Kemudian akan dicari nilai prediksinya berdasarkan nilai k yang ditentukan dari *KNN Regression* pada interval 3 sampai 20, disini akan menggunakan $k = 3$ terlebih dahulu. Hasil prediksi dapat diperoleh dari rata-rata nilai target yang diambil dari 3 data *training* dengan nilai terkecil. Berikut contoh perhitungan manual menggunakan rumus (2.1) pada 4 data prediksi dan hasil perhitungan seluruhnya dapat dilihat pada lampiran.

$$y'_{u1} = \frac{(5 + 11 + 17)}{3} = \frac{33}{3} = 11$$

$$y'_{u2} = \frac{(20 + 26 + 30)}{3} = \frac{76}{3} = 25,3$$

$$y'_{u3} = \frac{(13 + 24 + 25)}{3} = \frac{62}{3} = 20,6$$

$$y'_{u4} = \frac{(28 + 36 + 12)}{3} = \frac{76}{3} = 25,33$$

Dapat dilihat pada rumus di atas, data *testing* ke-1 memiliki hasil prediksi sebesar 11 sedangkan nilai targetnya adalah 10, dimana nilai selisihnya bernilai 1. Untuk melihat performa dari *fold* pertama, maka dibutuhkan nilai prediksi dari setiap data *testing*, dengan mengikuti langkah-langkah sebelumnya, berikut akan ditunjukkan dalam **Tabel 4.5** perbandingan nilai aktual dan prediksinya:

Tabel 4.5 Perbandingan Nilai Aktual dan Prediksi

Testing	Prediksi	Target
1	11	10
2	25,33	16
3	20,67	17

4	25,33	17
5	30	24

5. Menghitung Nilai MAD

Setelah mendapatkan nilai prediksi dari setiap data *testing*, maka akan dilakukan pengujian pada *fold* pertama dengan menggunakan salah satu dari *lost function* yaitu MAD atau MAE dengan rumus (2.3).

$$MAD = \frac{\sum_{i=1}^{X_{test}} |f(x_{test(i)}) - \hat{f}(x_{test(i)})|}{X_{test}}$$

$$MAD = \frac{|10 - 11| + |16 - 25,3| + |17 - 20,6| + \dots + |0 - 5,33|}{182} = 5,688$$

Setelah mendapatkan nilai MAD pada *fold* ke-1, maka akan dicari nilai MAD untuk *fold* ke-2 sampai ke-10, dengan mengikuti tahapan sebelumnya, maka hasil yang diperoleh akan diperlihatkan pada tabel berikut:

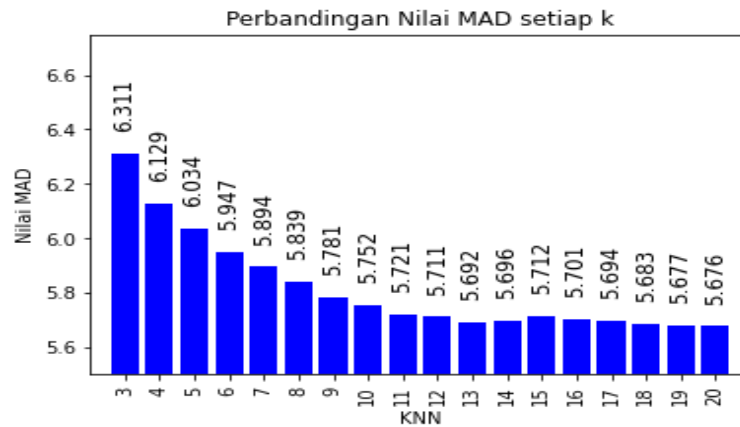
Tabel 4.6 Nilai MAD pada Setiap Fold

<i>n - Fold</i>	MAD
1	5,688
2	5,4
3	5,40
4	5,67
5	5,57
6	6,96
7	6,68
8	6,70
9	6,93
10	8,04
Rata - rata	6,31

Pada **Tabel 4.6** diperlihatkan, rata-rata *error* yang diperoleh sebesar 6,31, dengan *error* terkecil berada pada *fold* ke-3 dengan nilai 5,4, sedangkan *error* terbesarnya berada pada *fold* ke-10 dengan nilai 8,04.

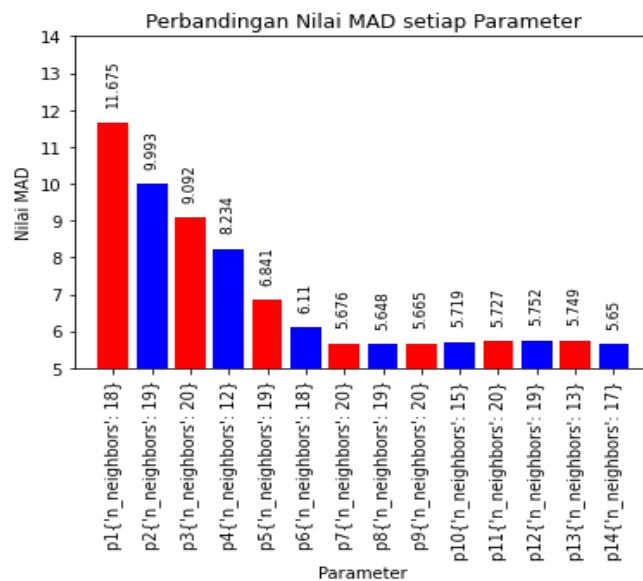
6. Pencarian Parameter

Setelah didapatkan nilai rata-rata MAD dari $k = 3$, maka akan dilihat nilai MAD untuk $k = 4, 5, 6, \dots, 20$. Hal ini bertujuan untuk mencari nilai k dengan *error* terkecil untuk penggunaan variabel input sebanyak 7 variabel.



Gambar 4.1 Diagram Rata-rata Nilai MAD pada Setiap k

Pada gambar 4.1 menunjukkan bahwasanya model dengan penggunaan 7 variabel input memberikan nilai MAD terkecil pada $k = 20$ sebesar 5,676. Antara nilai MAD yang terkecil dan terbesar, memiliki selisih kurang dari 1. Berikut disajikan plot rata-rata nilai MAD pada setiap k dari KNN *Regression*.



Gambar 4.2 Diagram Batang Rata-rata Nilai MAD pada Setiap Parameter

Gambar 4.2 menunjukkan bahwa, untuk data semua *game*, nilai MAD terkecil berada pada $k = 19$ sebesar 5,64 dengan penggunaan 8 variabel input.

4.4 Evaluasi Metode KNN Regression

Performa model dengan algoritma KNN Regression, ditinjau dari dua parameter yang mempengaruhinya, parameter k dari KNN Regression dan penggunaan jumlah variabel input. Model dengan parameter k terbaik antara 3 sampai 20, dan penggunaan jumlah variabel input antara 1 hingga 14 variabel pada semua data disajikan pada **Tabel 4.7**. Adapun pada **Tabel 4.7** menunjukkan semua data memiliki kesamaan antara satu sama lain pada parameter k dari KNN Regression lebih dari 18 dan penggunaan jumlah variabel lebih dari 7 input.

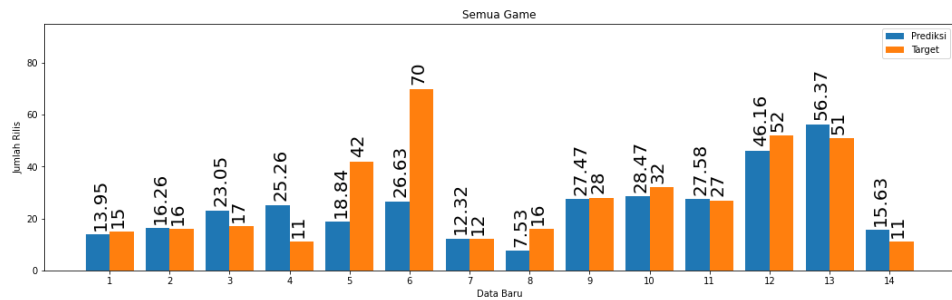
Tabel 4.7 Nilai MAD pada Setiap Data

Data	k	Jumlah Variabel Input	MAD
Semua <i>game</i>	19	8	5,64
Aksi	20	14	2,33
Petualangan	20	14	2,46
Kasual	20	14	2,23
Simulasi	20	14	1,60
Strategi	19	14	1,60
RPG	20	14	1,64

Adapun model dengan parameter terbaik yang telah didapatkan, akan diterapkan pada masing – masing data baru sebanyak 14 data atau selama 2 minggu berdasarkan penggunaan parameternya.

1. Data Semua *Game*

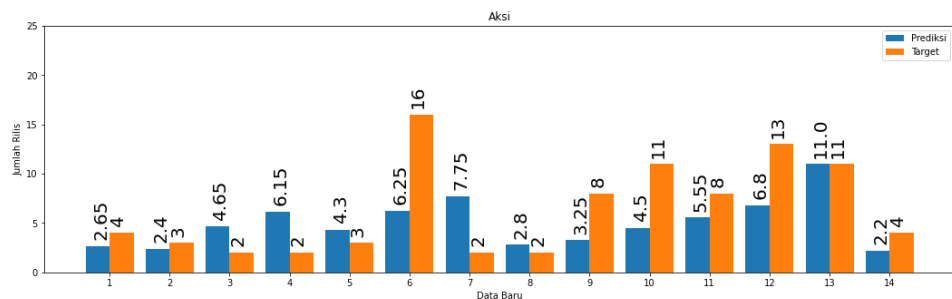
Pada data semua *game*, parameter k dari KNN *Regression* dengan nilai MAD terkecilnya berada pada $k = 19$ dan penggunaan 8 variabel input. **Gambar 4.3** merupakan perbandingan antara nilai prediksi dengan nilai target untuk data semua *game* dimana selisih terbesar memiliki nilai sebesar 43,37 pada data ke-6 dan selisih terkecilnya pada data ke-2 memiliki nilai sebesar 0,26.



Gambar 4.3 Plot Perbandingan Hasil Prediksi Data Semua *Game*

2. Data Aksi

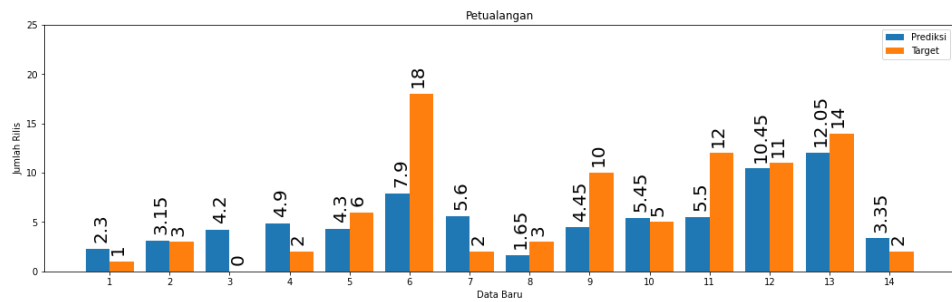
Untuk model dengan data bergenre aksi, parameter terbaiknya memiliki nilai MAD terkecil berada pada $k = 20$ dan penggunaan 14 variabel input. **Gambar 4.4** merupakan perbandingan antara nilai prediksi dengan nilai target untuk data bergenre aksi dimana selisih terbesar pada data ke-6 memiliki nilai sebesar 9,75 dan selisih terkecilnya pada data ke-13 memiliki nilai sebesar 0.



Gambar 4.4 Plot Perbandingan Hasil Prediksi Data Aksi

3. Data Petualangan

Pada model yang memiliki data bergenre petualangan, parameter dengan nilai MAD terkecil diperoleh dari $k = 20$ dan penggunaan 14 variabel input, perbandingan antara nilai prediksi dan targetnya di tunjukan pada gambar berikut.



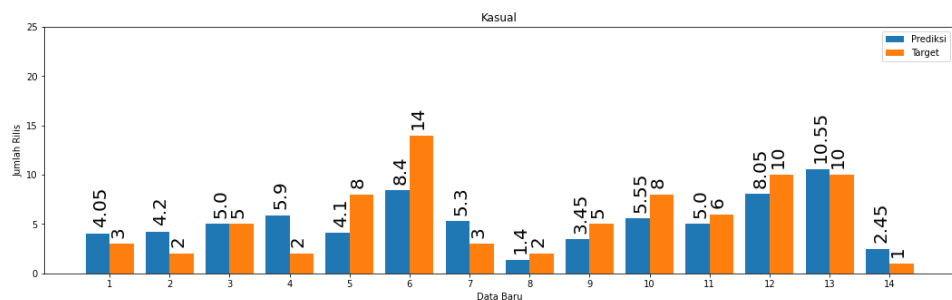
Gambar 4.5 Plot Perbandingan Hasil Prediksi Data Petualangan

Gambar 4.5 merupakan perbandingan antara nilai prediksi dengan nilai target untuk data bergenre petualangan yang dimana selisih terbesar memiliki nilai sebesar 10,1 pada data ke-6 dan selisih terkecilnya memiliki nilai sebesar 0,15 pada data ke-2.

4. Data Kasual

Model prediksi yang memiliki data bergenre kasual, parameter dengan nilai MAD terkecilnya berada pada $k = 20$ dan penggunaan 14 variabel input.

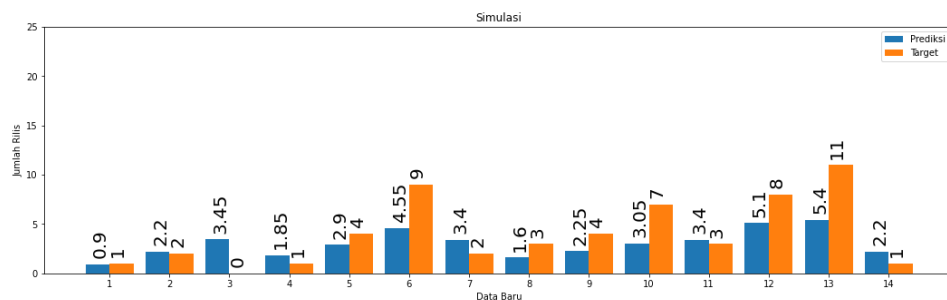
Gambar 4.6, selisih terbesar antara nilai target dan hasil prediksinya memiliki nilai sebesar 5,6 pada data ke-6 dan selisih terkecilnya memiliki nilai sebesar 0 pada data ke-3.



Gambar 4.6 Plot Perbandingan Hasil Prediksi Data Kasual

5. Data Simulasi

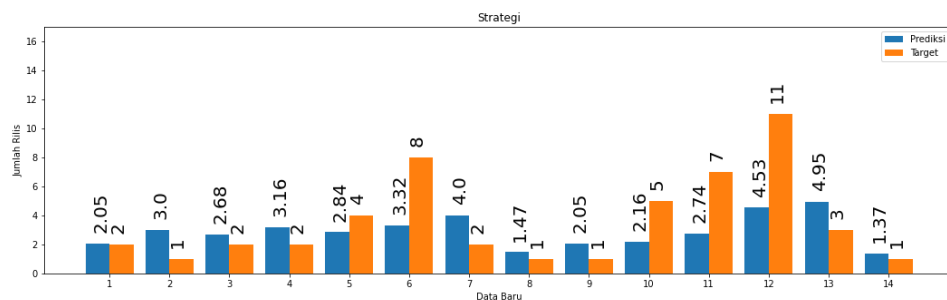
Adapun pada model yang memiliki data bergenre simulasi, parameter dengan nilai MAD terkecil berada pada $k = 20$ dan penggunaan 14 variabel input. **Gambar 4.7** menunjukkan perbandingan antara nilai prediksi dengan nilai target untuk data bergenre simulasi yang dimana selisih terbesar memiliki nilai sebesar 5,6 pada data ke-13 dan selisih terkecilnya memiliki nilai sebesar 0,1 pada data ke-1.



Gambar 4.7 Plot Perbandingan Hasil Prediksi Data Simulasi

6. Data Strategi

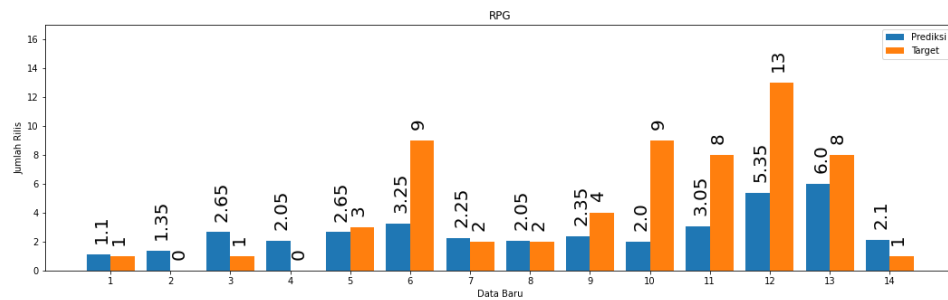
Pada model yang memiliki data bergenre strategi, parameter dengan nilai MAD terkecilnya berada pada $k = 19$ dan penggunaan 14 variabel input. **Gambar 4.8** menunjukkan perbandingan antara nilai prediksi dengan nilai target untuk data bergenre strategi yang dimana selisih terbesar memiliki nilai sebesar 6,47 pada data ke-12 dan selisih terkecilnya memiliki nilai sebesar 0,05 pada data ke-1.



Gambar 4.8 Plot Perbandingan Hasil Prediksi Data Strategi

7. Data RPG

Untuk model yang dibangun dengan data bergenre RPG, parameter yang memiliki nilai MAD terkecil, berada pada $k = 18$. **Gambar 4.9** menunjukkan perbandingan antara nilai prediksi dengan nilai target yang dimana selisih terbesar memiliki nilai sebesar 7,65 pada data ke-12 dan selisih terkecilnya memiliki nilai sebesar 0,05 pada data ke-8.



Gambar 4.9 Plot Perbandingan Hasil Prediksi Data RPG

Model dengan setiap jenis data, ketika diujikan terhadap data baru menggunakan perhitungan *error* yang sama pada pencarian parameter, terkadang hasil evaluasi dari nilai MAD-nya mengalami kenaikan maupun penurunan, hal ini disebabkan adanya pengaruh dari data baru dengan jumlah maupun nilai yang berbeda, untuk lebih jelasnya, nilai MAD pada setiap model dengan data yang berbeda, disajikan pada tabel berikut.

Tabel 4.8 Perbandingan Nilai MAD pada Setiap Data

Data	MAD	MAD pada Data Baru	Selisih
Semua <i>game</i>	5,64	8,39	2,75
Aksi	2,33	3,43	1,1
Petualangan	2,46	2,98	0,52

Kasual	2,23	2,04	0,19
Simulasi	1,60	2,08	0,48
Strategi	1,60	2,56	0,96

4.5 Prediksi dalam Islam

Telah dibahas pada bab sebelumnya, prediksi jumlah rilis *game* berkaitan dengan data pada waktu lampau yang mana dalam surah Yunus ayat 10 berbunyi:

هُوَ الَّذِي جَعَلَ الشَّمْسُ ضِيَاءً وَالْقَمَرَ نُورًا وَقَدَرَهُ مَنَازِلَ لِتَعْلَمُوا عَدَدَ السِّنِينَ وَالْحِسَابَ ۗ مَا خَلَقَ اللَّهُ ذَلِكَ إِلَّا بِالْحَقِّ ۗ يُفَصِّلُ الْآيَاتِ لِقَوْمٍ يَعْلَمُونَ

“Dialah yang menjadikan matahari bersinar dan bulan bercahaya dan ditetapkan-Nya manzilah-manzilah (tempat-tempat) bagi perjalanan bulan itu, supaya kamu mengetahui bilangan tahun dan perhitungan (waktu). Allah tidak menciptakan yang demikian itu melainkan dengan hak. Dia menjelaskan tanda-tanda (kebesaran-Nya) kepada orang-orang yang mengetahui”

Penjelasan ayat di atas berdasarkan tafsir Quraish Shihab adalah bahwasanya Allah menciptakan bumi-langit dengan matahari yang memberikan pancaran sinar di saat siang dan bulan yang memantulkan cahaya matahari di saat malam. Allah menentukan wilayah yang berbeda sebagai peredaran bulan sehingga pancaran cahaya bulan akan mengikuti wilayah tersebut, hal ini dimaksudkan agar manusia dapat mengetahui dan memprediksi waktu serta penanggalan tahun. Dibandingkan dengan tempat beredarnya, bulan memiliki perbedaan dengan matahari, sehingga menghasilkan pola bulan yang berbeda. Dari penjelasan ini, penetapan kalender komariah yang berdasarkan penglihatan bulan dan angkasa, waktu yang ditentukan adalah 29 hari, 12 jam dan 44 menit. Sama halnya dengan prediksi jumlah rilis *game* untuk kedepannya, yang berkaitan dengan perhitungan jumlah rilis berdasarkan waktu yang lalu.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan pembahasan pada bab 4, dapat ditarik kesimpulan bahwa metode KNN *Regression* yang dibuat kedalam model prediksi jumlah rilis *game* pada platform Steam, data dengan rentang waktu rilis dari tahun 2018 hingga 2022, memiliki kesamaan pada pencarian parameter berdasarkan nilai MAD terkecil, dimana parameter k dari KNN *Regression* memiliki nilai lebih dari 18 dan penggunaan jumlah variabel input lebih dari 8.

Adapun hasil evaluasi berdasarkan perhitungan nilai MAD, model dengan data semua *game* memiliki nilai tertinggi sebesar 5,64 pada $k = 19$ dari KNN *Regression* dan penggunaan 8 variabel input, sedangkan model dengan data bergenre simulasi dan strategi, memiliki nilai terendah yaitu 1,60 pada $k = 20$ dan $k = 19$ serta penggunaan 14 variabel input. Hasil prediksi dengan melakukan pengujian pada data baru, model yang dibuat dengan data semua *game*, memiliki nilai selisih terbesar antara hasil prediksi dengan nilai target yaitu sebesar 43,7, adapun model dengan data bergenre aksi dan kasual, memiliki nilai selisih terkecil yaitu sebesar 0. Berdasarkan hasil evaluasi yang diperoleh, metode KNN *Regression* dapat diterapkan dalam memprediksi jumlah *game* yang rilis pada platform Steam.

5.2 Saran

Pada penelitian selanjutnya, dapat dilakukan dengan mencari parameter pada jangkauan yang lebih luas dengan cara memaksimalkan parameter k dari

KNN *Regression* dan penggunaan jumlah variabel input. Pengujian dengan mengubah pembagian data *training* dan data *testing* dapat mempengaruhi hasil evaluasi.

DAFTAR PUSTAKA

- Andini, E., Faisal, M. R., Herteno, R., Nugroho, R., Abadi, F., & Muliadi, M. (2022). Peningkatan Kinerja Prediksi Cacat Software Dengan Hyperparameter Tuning Pada Algoritma Klasifikasi Deep Forest. *Jurnal Mnemonic*, 5, 119–127. <https://doi.org/10.36040/mnemonic.v5i2.4793>
- Anggie, A. (2018). Implementasi Metode K-Nearest Neighbor Dan Regresi Linear Dalam Prediksi Harga Emas. *Pemodelan Arsitektur Sistem Informasi Perizinan Menggunakan Kerangka Kerja Togaf Adm*, 4(1), 113. <http://e-journal.janabadra.ac.id/>
- Ansori, M. (2020). *Metode Penelitian Kuantitatif Edisi 2*. Airlangga University Press. <https://books.google.co.id/books?id=rKbJDwAAQBAJ>
- Drajana, I. C. R. (2018). Prediksi Jumlah Produksi Coconut Oil Menggunakan k-Nearest Neighbor dan Backward Elimination. *Tecnoscienza*, 13(1), 51–64.
- Fauzi, M. (2009). *Metode Penelitian Kuantitatif Sebuah Pengantar* (M. N. Ichwan (ed.)). Walisongo Press.
- Grewal, B., Lin, D., & Bezemer, C.-P. (2022). An Empirical Study of Delayed Games on Steam. *Arxiv*.
- Hamdi, A., Indriani, F., & Muliadi, M. (2019). Metode Timeseries K-Nearest Neighbor Regression Dalam Prediksi Barang Keluar Pada Gudang Pt Putra Prenuer Banjarbaru. *Seminar Nasional Ilmu Komputer (SOLITER)*, 2, 37–45. <https://soliter.ulm.ac.id/index.php/soliter/article/view/51>
- Hasan, I. (2006). *Analisis data penelitian dengan statistik* (2nd ed.). Bumi Aksara.
- Imron, I. (2019). Analisa Pengaruh Kualitas Produk Terhadap Kepuasan Konsumen Menggunakan Metode Kuantitatif Pada CV. Meubele Berkah Tangerang. *Indonesian Journal on Software Engineering (IJSE)*, 5(1), 19–28. <https://doi.org/10.31294/ijse.v5i1.5861>
- Juwiantho, H. B. R. I. H. (2022). Pemanfaatan text summarization dengan Support Vector Machine dan K-nearest neighbor pada analisis sentimen untuk mempermudah pengguna membaca review game STEAM. *Jurnal Infra, Vol 10, No 1 (2022)*, 31–36. <https://publication.petra.ac.id/index.php/teknik-informatika/article/view/11895/10489>
- Kapp, K. M., Valtchanov, D., & Pastore, R. (2020). Enhancing motivation in workplace training with casual games: a twelve month field study of retail employees. *Educational Technology Research and Development*, 68(5), 2263–2284. <https://doi.org/10.1007/s11423-020-09769-2>
- Li, X., & Zhang, B. (2020). *A preliminary network analysis on steam game tags: another way of understanding game genres*. <https://doi.org/10.1145/3377290.3377300>
- Mulyati, S., Husein, S. M., & Ramdhan, R. (2020). Rancang Bangun Aplikasi Data

- Mining Prediksi Kelulusan Ujian Nasional Menggunakan Algoritma (Knn) K-nearest Neighbor dengan Metode Euclidean Distance pada SMPN 2 Pagedangan. *Jurnal Informatika*, 4(1), 65–73. <https://doi.org/10.31000/jika.v4i1.2288>
- Muslich, M. (2009). *Metode Pengambilan Keputusan Kuantitatif*. Sinar Grafika.
- Nurfauzan, D., & Fatimah, T. (2022). Implementasi Algoritma K-Nearest Neighbors Regression Dalam Memprediksi Harga Saham. In *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia*. <https://senafti.budiluhur.ac.id/index.php/>
- Pérez-Marcos, J., Sánchez-Moreno, D., Batista, V. L., & Muñoz, M. D. (2019). *Estimated Rating Based on Hours Played for Video Game Recommendation BT - Distributed Computing and Artificial Intelligence, Special Sessions, 15th International Conference* (S. Rodríguez, J. Prieto, P. Faria, S. Klos, A. Fernández, S. Mazuelas, M. D. Jiménez-López, M. N. Moreno, & E. M. Navarro (eds.); pp. 300–307). Springer International Publishing.
- Prado, R., & West, M. (2010). *Time series: modeling, computation, and inference*. Chapman and Hall/CRC.
- Qaffas, A. A. (2020). An operational study of video games genres. *International Journal of Interactive Mobile Technologies*, 14(15), 175–194. <https://doi.org/10.3991/IJIM.V14I15.16691>
- Saefudin, A., Notodiputro, K. A., Alamudi, A., & Sadik, K. (2009). *Statistika Dasar*. Grasindo.
- Santosa, B. (2007). *Data mining : Teknik pemanfaatan data untuk keperluan bisnis*. Garah Ilmu.
- Schefler, W. C. (1987). *Statistika untuk Biologi, Farmasi, Kedokteran dan Ilmu yang Bertautan*. ITB Bandung.
- Seruni, D. S., Furqon, M. T., & Wihandika, R. C. (2020). *Sistem Prediksi Pertumbuhan Jumlah Penduduk Kota Malang menggunakan Metode K-Nearest Neighbor Regression* (Vol. 4, Issue 4). <http://j-ptiik.ub.ac.id>
- Shihab, M. Q. (2021). *Al-Quran dan Maknanya* (3rd ed.). Lentera Hati.
- Subagyo, P. (2010). *Statistika Terapan* (2nd ed.). BPFE - Yogyakarta.
- Sugiyono. (2018). *Metode Penelitian Evaluasi* (1st ed.). Alfa Beta.
- Tanuwijaya, J., & Hansun, S. (2019). LQ45 Stock Index Prediction using k-Nearest Neighbors Regression. In *International Journal of Recent Technology and Engineering* (Vol. 8). <https://doi.org/10.35940/ijrte.C4663.098319>
- Yuwantoro, M., Mahmud, I., & Murdiansyah Danang Triantoro, T. U. (2019). Prediksi Harga Beras Premium dengan Metode Algoritma K-Nearest Neighbor. *E-Proceeding of Engineering*, 7(1), 2714–2724.

LAMPIRAN

Lampiran 1 Data Steam

Data Rilis Perhari 2018 – 2022

<https://bit.ly/3CIDS7>

Hasil Jarak Data *Test* ke-1 pada *Fold* ke-1

<https://bit.ly/3phKgOu>

Hasil Prediksi *Fold* ke-1

<https://bit.ly/46gGrd8>

Data *Time Series* setiap Genre

<https://bit.ly/3Poe8Ud>

Lampiran 2 Source Code

```
import openpyxl
from openpyxl import Workbook
from openpyxl.styles import Font
import pandas as pd

import numpy as np
from sklearn.model_selection import KFold

from sklearn.metrics import mean_absolute_error as mae
from sklearn.neighbors import KNeighborsRegressor
from sklearn.model_selection import GridSearchCV
import matplotlib.pyplot as plt
```

Data

```
berkas= 'D:\GData\AllGame.xlsx'
data = pd.read_excel(berkas)
```

Pencarian Parameter

```
bk = []
bs = []
bp = []

p = ['p1','p2','p3','p4','p5','p6','p7','p8','p9','p10','p11','p12','p13','p14']
for i in p:
    data = pd.read_excel(berkas, sheet_name = i)
    X = data.iloc[:round(len(data)-14), :-1].values
    y = data.iloc[:round(len(data)-14), -1:].values
    knn = KNeighborsRegressor(metric='euclidean')

    param_grid = {'n_neighbors' : np.arange(3,21)}
    knn_gscv = GridSearchCV(knn, param_grid, cv=10, scoring= "neg_mean_absolute_error")
    knn_gscv.fit(X,y)
    print(f'nilai MAD terendah pada {i} adalah',-knn_gscv.best_score_, f'dengan {knn_gscv.best_params_}')

    bk.append(i)
    bs.append(-knn_gscv.best_score_ )
    bp.append(knn_gscv.best_params_)
```

Pengujian Data Baru

```
df = pd.DataFrame(zip(bk,bs,bp), columns=['Parameter','Nilai MAD','KNN'])
dfs = df.sort_values(by=['Nilai MAD'])

hp = dfs['Parameter'].loc[dfs.index[0]]
data = pd.read_excel(berkas, sheet_name = hp)

X = data.iloc[:round(len(data)-14), :-1].values
y = data.iloc[:round(len(data)-14), -1:].values
X_test = data.iloc[round(len(data)-14):, :-1].values
y_test = data.iloc[round(len(data)-14):, -1:].values

knn = KNeighborsRegressor()
param_grid = {'n_neighbors' : np.arange(1,21)}
knn_gscv = GridSearchCV(knn, param_grid, cv=10, scoring="neg_mean_absolute_error" )
knn_gscv.fit(X,y)
prediksi = knn_gscv.predict(X_test)
pred = knn_gscv.best_estimator_.predict(X_test)
print('Nilai MAD = ',mae(pred, y_test))
gh = pd.DataFrame(zip(pred,y_test,abs(pred-y_test)), columns=['Prediksi','Target','Selisih'])
gh
```

Plot

```
x1=np.arange(1,len(X_test)+1)
x2=np.arange(1,len(X_test)+1)
y1=pred
y2=y_test
plt.plot(x1, y1, label="Prediksi",color='blue', linestyle='dashed', linewidth = 3,
        marker='o', markerfacecolor='blue', markersize=12)
plt.plot(x2, y2, label="Target",
        marker='o', markerfacecolor='red', markersize=12)

# naming the x axis
plt.xlabel('Waktu Rilis')
# naming the y axis
plt.ylabel('Jumlah Game')
# giving a title to my graph
plt.title('Semua Game')

# show a legend on the plot
plt.legend()

plt.xlim(0,15)

plt.show()
```

RIWAYAT HIDUP



Penulis bernama Sayid Syeh Yusuf Al-Balgaist merupakan laki-laki kelahiran 4 Agustus 1997 di daerah kota Lamongan, Jawa Timur. Sejak memulai pendidikan formal pada jenjang sekolah dasar, penulis mulai berdomisili di kota Sampit, Kabupaten Kotawaringin Timur, Kalimantan Tengah.

Tentunya, penulis menempuh pendidikan sekolah dasar di SDN 4 Sawahan, Sampit dan melanjutkan SMP serta SMA di pesantren plus Al-Mahadul Islami, Bangil, Pasuruan, Jawa Timur. Kemudian pada tahun 2016, penulis melanjutkan ke jenjang lebih tinggi di Universitas Islam Negeri Maulana Malik Ibrahim Malang, pada Program Studi Matematika.



BUKTI KONSULTASI SKRIPSI

Nama : Sayid Syeh Yusuf Al - Balgaist
NIM : 16610112
Fakultas / Jurusan : Sains dan Teknologi / Matematika
Judul Skripsi : Prediksi Jumlah Rilis *Game* Pada Platform Steam dengan Menggunakan Algoritma *K-Nearest Neighbor Regression*
Pembimbing I : Hisyam Fahmi, M. Kom.
Pembimbing II : Mohammad Nafie Jauhari, M.Si.

No	Tanggal	Hal	Tanda Tangan
1.	16 Januari 2023	Konsultasi BAB 1, 2 dan 3	1.
2.	20 Januari 2023	Revisi BAB 1, 2 dan 3	2.
3.	26 Januari 2023	ACC Kajian Agama	3.
4.	27 Januari 2023	Konsultasi BAB 3	4.
5.	02 Februari 2023	ACC BAB 1, 2 dan 3	5.
6.	24 Februari 2023	Konsultasi PPT Sempro	6.
7.	03 Maret 2023	Konsultasi BAB 4 dan 5	7.
8.	10 Maret 2023	Revisi BAB 4 dan 5	8.
9.	17 Maret 2023	ACC BAB 4 dan 5	9.
10.	19 Mei 2023	Konsultasi Persiapan Semhas	10.
11.	19 Mei 2023	ACC Kajian Agama	11.
12.	7 Juni 2023	Revisi BAB 4,5 Dan Abstrak	12.
13.	16 Juni 2023	ACC Sidang	13.
14.	21 Juni 2023	ACC Keseluruhan	14.

Mengetahui,
Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc
NIP.197411292000122005

