

# Analisis Komparatif ARIMA dan Prophet dengan Studi Kasus Dataset Pendaftaran Mahasiswa Baru

<http://dx.doi.org/10.28932/jutisi.v6i2.2676>

Cato Chandra #<sup>1</sup>, Setia Budi ✉<sup>\*2</sup>

*#\*Magister Ilmu Komputer, Universitas Kristen Maranatha  
Jl. Prof.Drg. Surya Sumantri, MPH 65. Bandung*

<sup>1</sup>cato.chandra@maranatha.edu

<sup>2</sup>setia.budi@it.maranatha.edu

**Abstract** — This research presents all studies, methodologies, and results about testing the accuracy of predictions on new student marketing data by region using the Prophet and Autoregressive Integrated Moving Average (ARIMA) methods. The dataset selected for this study uses 26 years of actual data that has an annual interval. The data was prepared for time series forecasting analysis, therefore, several numbers of data preprocessing were applied such as log transformation and resampling. To get efficient variables, the best variables will be sought to improve the accuracy of predictions. Both models will conduct training and test data to produce values that can be compared using the metric regression model. Based on the training conducted, Prophet has better performance than ARIMA.

**Keywords**—ARIMA; forecasting; FB Prophet; time series data analysis.

## I. PENDAHULUAN

Kesuksesan tidak dapat dicapai apabila tidak ada perencanaan yang matang. Dengan perencanaan yang matang maka akan menghasilkan keputusan yang berkualitas [1] dan keputusan yang baik adalah keputusan yang didasarkan atas pertimbangan apa yang akan terjadi pada waktu keputusan itu dilaksanakan. Organisasi tidak akan bisa berkembang apabila tidak memiliki strategi yang tepat. Prinsip yang sama juga berlaku dalam bidang pendidikan, setiap keputusan yang dihadapi harus didasarkan pada data yang dapat menghasilkan pola pikir dan probabilitas perkiraan terkait apa yang akan dihadapi di masa mendatang. Peran serta data terhadap pengambilan keputusan menjadi pondasi utama dalam perencanaan strategis. Universitas Kristen Maranatha (UK. Maranatha) sebagai institusi perguruan tinggi memerlukan perencanaan strategis yang berfungsi untuk meningkatkan eskalasi mahasiswa sehingga berkembang kearah yang lebih baik.

### A. Latar Belakang

UK. Maranatha sebagai universitas swasta yang pendanaan utamanya bersumber dari uang kuliah (*tuition fee*)

mahasiswa, maka angka pendaftaran mahasiswa baru sangatlah krusial. Untuk menarik mahasiswa baru, dibutuhkan perencanaan pemasaran yang tepat. Perencanaan pemasaran tidak hanya berdasarkan insting dalam menentukan wilayah mana yang sedang tren, melainkan dibutuhkan data yang memadai.

Dengan dukungan data yang memadai maka perencanaan pemasaran akan semakin kuat karena didukung oleh sistem pengambilan keputusan yang baik. Penentuan lokasi pemasaran adalah salah satu dari poin penting dalam perencanaan pemasaran. Untuk mendapatkan lokasi pemasaran yang tepat, maka peran teknologi informasi diperlukan. Dengan adanya data historis terkait pendaftaran mahasiswa baru, maka didapatkan pula data lokasi sekolah tempat asal mahasiswa yang mendaftar.

Data tersebut dapat diolah dan dianalisis dengan menggunakan peramalan deret waktu sebagai salah satu alat bantu untuk melakukan perhitungan dengan pendekatan kuantitatif agar mendapatkan hasil probabilitas jumlah pendaftar di masa depan [2].

Peramalan deret waktu sering kali digunakan untuk membantu kesuksesan dalam perusahaan sebagai alat keputusan strategi manajerial [3]. UK. Maranatha memerlukan *strategic planning* yang berfungsi untuk meningkatkan eskalasi mahasiswa. Dengan adanya peramalan, diharapkan dapat memberikan gambaran strategi dalam pengambilan keputusan untuk menentukan wilayah pemasaran dimasa yang akan datang.

Ada beberapa metode yang ditawarkan dalam peramalan deret waktu [4], yaitu: *Box-Jenkins ARIMA models*, *Box-Jenkins Multivariate models*, dan *Holt-Winters Exponential Smoothing*. Peramalan dengan menggunakan metode ARIMA populer digunakan dikalangan *data scientist* [5] dan umum diterapkan untuk melakukan peramalan data deret waktu. Akan tetapi untuk menghasilkan peramalan yang baik tidaklah mudah, terdapat dua kendala besar yang dihadapi oleh para peramal bisnis [6], yaitu: teknik peramalan secara otomatis terbilang sulit untuk di *tune* karena sering kali tidak fleksibel untuk memasukkan asumsi

yang dibutuhkan atau dapat disebut juga *heuristic*, kedua adalah analisis yang bertanggung jawab untuk tugas ilmu data di seluruh organisasi biasanya memiliki pengetahuan domain yang mendalam terkait produk atau layanan tertentu yang mereka dukung, tetapi sering tidak memiliki pelatihan dalam peramalan deret waktu.

Tidak semua masalah peramalan dapat diselesaikan dengan prosedur yang sama, oleh sebab itu dengan adanya metode yang ditawarkan oleh *Facebook Research*, yaitu *Prophet*. *Prophet* dioptimalkan untuk mengatasi masalah peramalan bisnis yang dihadapi. *Prophet* dapat menggantikan *forecast package* yang ada, sebab *Prophet* memiliki dua kelebihan dari *forecast package* lainnya [6], yaitu:

- Penggunaan *Prophet* jauh lebih mudah untuk membuat perkiraan yang wajar dan akurat karena banyak dari metode peramalan yang ditawarkan seperti ARIMA, *Exponential Smoothing*, dan metode lainnya memiliki parameter *tuning* masing-masing yang dapat menyebabkan hasil yang buruk apabila analisis tidak berpengalaman dalam mengatur parameternya, sebab untuk menentukan parameter yang efisien itu banyak pilihan cara.
- *Prophet* dapat disesuaikan dengan cara *intuitive* untuk analisis yang belum berpengalaman. Maksudnya adalah *Smoothing Parameter* untuk *seasonal* dapat diatur seberapa dekat dengan siklus *historical data*-nya.

Dengan kemudahan yang ditawarkan oleh metode *Prophet* tersebut, maka analisis pemula dapat beradaptasi dengan mudah dan menghasilkan ramalan yang reliabel dan akurat.

Untuk mengukur keakuratan dari hasil ramalan *Prophet*, ada beberapa penelitian yang mendukungnya, salah satunya adalah penelitian yang dilakukan oleh Isil Yenidogan dan Aykut Cayir tentang penerapan metode ARIMA dan *Prophet* diterapkan pada interval data per jam dari data *Bitcoin* [7] yang mengatakan bahwa teknik peramalan deret waktu *Prophet* mampu berperforma 245.09 (dalam nilai RMSE) lebih baik dari pada teknik peramalan deret waktu ARIMA yang memiliki nilai 593.80 (dalam nilai RMSE) berdasarkan nilai dari *testing data* yang dilakukan dari bulan May 2016 sampai dengan Januari 2018.

Dari hasil tersebut, penelitian ini akan menerapkan dua metode yaitu ARIMA dan *Prophet* ke dalam data yang digunakan di UK. Maranatha, yaitu data registrasi mahasiswa. Data tersebut akan diolah untuk menghasilkan seberapa besar performa deret waktu yang dihasilkan oleh metode ARIMA dan *Prophet*

### B. Rumusan Masalah

Berdasarkan latar belakang tersebut, maka disusunlah rumusan masalah yang sedang dihadapi oleh tim pemasaran UK. Maranatha, sebagai berikut:

- Metode peramalan (*forecasting*) apakah yang paling tepat untuk diterapkan dalam membantu tim pemasaran universitas?

- Bagaimana mekanisme untuk mengukur performa dari suatu metode peramalan dalam kaitannya dengan data penerimaan mahasiswa baru?

Melihat pada masalah-masalah yang ada pada tim pemasaran UK. Maranatha, maka dilakukan penelitian tentang analisis data dengan menggunakan teknik peramalan (*forecasting*) dengan menggunakan metode peramalan *Facebook Prophet (Prophet)* dan membandingkan metode tersebut dengan metode *AutoRegressive Integrated Moving Average (ARIMA)*.

### C. Hipotesis

Dari penelitian yang dilakukan oleh Isil Yenidogan dan Aykut Cayir [7], penerapan metode ARIMA dan *Prophet* diterapkan pada interval data per jam dari data *Bitcoin* menghasilkan tingkat performa *Prophet* yang lebih baik dari pada ARIMA. Dalam penelitian ini, akan menerapkan perbandingan performa dari metode *Prophet* dan ARIMA dengan menggunakan dataset pendaftaran mahasiswa baru yang memiliki interval data tahunan (*annual*). Dari hal tersebut dapat dibuat hipotesis sebagai berikut:

- $H_0$ , artinya ramalan yang dihasilkan oleh ARIMA dan *Prophet* terhadap data mahasiswa yang mendaftar ke UK. Maranatha mendekati nilai nol.
- $H_1$ , artinya ramalan yang dihasilkan oleh *Prophet* memiliki performa yang lebih baik dari ARIMA.

### D. Tujuan Penelitian

Penelitian ini bertujuan untuk membantu Direktorat Pemasaran UK. Maranatha sebagai penunjang data pengambilan keputusan. Dengan melakukan teknik peramalan deret waktu terhadap data pendaftaran calon mahasiswa baru dengan harapan mendapatkan hasil ramalan untuk setiap kota dan provinsi untuk menjadi bahan pertimbangan dalam menentukan strategi pemasaran.

*Prophet* sebagai solusi yang ditawarkan untuk metode peramalan utama yang digunakan dalam penelitian ini. Akan tetapi agar mendapatkan hasil yang optimum maka akan dilakukan perbandingan analisis data terhadap *Prophet* dengan menggunakan ARIMA. Adapun harapan lain yang ingin dicapai dalam penelitian ini, yaitu ingin membuktikan bahwa *Prophet* akan mendapatkan performa yang lebih baik dibandingkan dengan ARIMA.

### E. Ruang Lingkup

Ruang lingkup penelitian terbatas pada Direktorat Pemasaran UK. Maranatha dan direktorat lain yang memiliki hubungan erat dengan lainnya seperti, Direktorat Informasi dan Direktorat Akademik. Dengan batasan-batasan yang diberikan oleh setiap direktorat, maka pengembangan penelitian berdasarkan batasan tersebut.

Data yang akan digunakan sebagai studi kasus dalam penelitian ini, yaitu data registrasi mahasiswa UK. Maranatha tahun 1993 sampai dengan 2019.

Untuk teknik peramalan yang akan dipergunakan dalam penelitian ini adalah menggunakan model peramalan deret

waktu. Dalam peramalan deret waktu ada beberapa teknik yang dapat dipergunakan, teknik utama yang ingin dipergunakan dalam penelitian ini adalah menggunakan *Prophet*. Selain dari teknik utama yang diajukan, adapun teknik lain yang digunakan sebagai teknik pembanding, yaitu ARIMA. Harapan yang akan dihasilkan dari metode pembanding ini akan dipergunakan untuk membandingkan keakuratan data ramalan yang dihasilkan dari metode utama.

Penelitian ini menggunakan *Jupyter Notebook* dengan bahasa pemrograman *Python* sebagai media pemrograman dan juga media pemrosesan data hingga penyampaian hasil praktik terhadap penelitian yang dilakukan

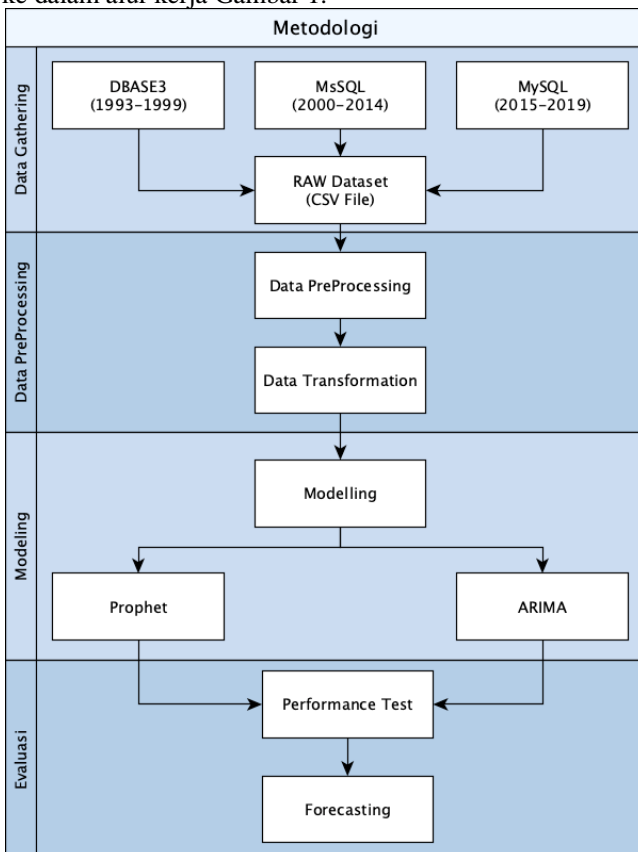
## II. MODEL KEBUTUHAN DAN ANALISIS

### A. Konsep Dasar Operasional

Konsep dasar operasional yang terkait dengan penelitian ini adalah:

- Data yang digunakan adalah data deret waktu, yaitu data yang dikumpulkan dari untaian waktu tertentu dan menggambarkan perkembangan suatu kegiatan yang berlangsung.
- Data sekunder adalah data yang didapat dari lembaga atau instansi tertentu yang mendukung tujuan penelitian.

Kerangka pemikiran dalam penelitian ini dapat digambarkan ke dalam alur kerja Gambar 1.



Gambar 1. Metodologi

Dari Gambar 1 dapat dibagi menjadi empat tahapan yaitu tahapan pengumpulan data (*Data Gathering*), pengolahan data (*Data Preprocessing*), penerapan metode (*Modeling*), dan evaluasi data (*Data Testing*).

### B. Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data historis registrasi mahasiswa baru dari tahun 1993 sampai 2019 sebanyak 159.770 data pendaftar yang dikelompokkan berdasarkan tahun dan nama kota dimana setiap kota memiliki satu dataset dengan 26 data observasi. Sedangkan untuk melakukan permodelan peramalan, minimal observasi yang dapat dilakukan permodelan peramalan sebanyak 16 data observasi dengan keakuratan hasil ramalan masih dapat diterima [8].

### C. Metode Analisis Data

Dalam penelitian ini digunakan adalah metode analisis yang terdiri dari analisis deskriptif dan analisis inferensial yang berupa analisis deret waktu. Namun, tetap dapat melakukan ekstraksi data dengan teknik analisis yang berbeda-beda tergantung kebutuhan. Dengan menggunakan analisis konten maka dapat membantu memahami keseluruhan tema yang ada dalam data kuantitatif yang dimiliki. Dengan metode ini dapat membantu mengurai data tekstual dan dapat menemukan rangkaian data yang umum digunakan.

## III. LANDASAN TEORI

### A. Peramalan (Forecasting)

*Forecasting* dapat dilakukan dengan adanya data historis yang dikalkulasikan dengan metode tertentu untuk memproyeksikan sebuah prediksi yang akan terjadi. *Forecasting* adalah masukan dasar dari proses pengambilan keputusan manajemen karena hal tersebut sebagai penyedia informasi yang dibutuhkan dimasa depan [9].

Tujuan dari *forecasting* adalah untuk menentukan keakuratan dan kekuatan yang diperlukan oleh teknik yang dipilih sehingga memutuskan perkiraan mana yang memungkinkan memasuki area bisnis dari ukuran pasar yang ada [10]. Sedangkan peramalan atau *forecasting* memiliki tujuan untuk mengkaji kebijakan perusahaan yang berlaku saat ini dan di masa lalu serta melihat sejauh mana pengaruh di masa datang.

### B. Analisis Deret Waktu dan Proyeksinya

*Time Series Analysis* adalah teknik statistik yang digunakan ketika data memiliki data historis untuk produk atau garis produk yang tersedia dan memiliki hubungan dan tren yang jelas dan relatif stabil.

Salah satu prinsip dasar peramalan deret waktu adalah ketika data historis tersedia, maka pelaku peramalan harus menggunakan data kinerja masa lalu tersebut untuk mendapatkan “pembacaan” sampai tingkat saat ini, seberapa cepat perubahan meningkat atau menurun. Tingkat perubahan tersebut merupakan dasar dari peramalan. Setelah

mengetahui tingkatan tersebut, maka dapat dikembangkan proyeksinya dengan menggunakan teknik matematika.

Hal tersebut tidaklah mudah, kendala yang dihadapi dalam membuat proyeksi dari data mentah adalah tingkat dan tren yang tidak jelas, karena dicampur dengan variasi musiman, misalnya ada faktor distorsi seperti efek peningkatan drastis atau efek penurunan drastis akibat event tertentu. Dengan adanya kendala tersebut, data mentah harus diolah terlebih dahulu sebelum digunakan, dan ini disebut sebagai analisis deret waktu.

Deret waktu adalah sebuah set titik kronologis yang diurutkan dari data mentah, misalnya penjualan produk tertentu, per bulan, selama beberapa tahun. Analisis deret waktu membantu mengidentifikasi dan menjelaskan setiap urutan atau variasi sistematis dalam rangkaian data secara musiman (dikenal sebagai *seasonal data*). Sedangkan pola siklus yang mengulang akan menjadi sebuah tren dalam data.

Setelah analisis selesai, maka akan dapat menghasilkan sebuah proyeksi perkiraan dimasa depan. Perlu diingat bahwa meskipun telah berhasil memisahkan analisis dari proyeksi, sebagian besar dari teknik peramalan statistik adalah menggabungkan kedua fungsi dalam satu operasi.

#### C. Augmented Dickey Fuller (ADF)

Uji stasioner dapat menggunakan *Augmented Dickey Fuller* (ADF), dimana pengujian ini menentukan apakah ada data deret waktu yang mengandung akar unit (*unit root*). Pengujian dapat dilakukan dengan menguji hipotesis dalam persamaan regresi sebagai berikut:

$$Y_t = \varphi Y_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta Y_{t-j} + \theta_0 + a_t \quad (1)$$

Pada persamaan (1), memiliki hipotesis sebagai berikut:

- $H_0: \varphi = 1$  berarti data mengandung akar unit yang berarti tidak stasioner
  - $H_0: |\varphi| < 1$  berarti data tidak mengandung akar unit
- Atau

$$\Delta Y_t = \delta Y_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta Y_{t-j} + \theta_0 + a_t \quad (2)$$

Pada persamaan (2), memiliki hipotesis sebagai berikut:

- $H_0: \delta = 1$  berarti data mengandung akar unit yang berarti tidak stasioner
- $H_0: \delta < 1$  berarti data tidak mengandung akar unit

Hipotesis nol ditolak jika nilai *Test Statistic* memiliki nilai kurang (lebih negatif) daripada *critical value*. Secara umum, *p-value* kurang dari 0.05 dapat disebut juga stasioner [11].

#### D. Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

KPSS digunakan untuk mencari deret waktu stasioner diantara rata-rata atau *linear trend*, atau tidak stasioner karena akar unit. KPSS berdasarkan dari regresi linear yang dipecah menjadi tiga bagian, yaitu: *deterministic trend* ( $\beta_t$ ), *random walk* ( $r_t$ ), dan *stationary error* ( $\varepsilon_t$ ), dengan rumus regresi sebagai berikut:

$$x_t = r_t + \beta_t + \varepsilon_t \quad (3)$$

Jika data stasioner, maka akan mempunyai elemen tetap untuk intersepsi atau deret akan menjadi stasioner sekitar level tetap [12]. Atau, jika *Test Statistic* lebih besar dari *p-value* maka data *trend* tersebut stasioner [13]

#### E. Autoregressive integrated moving average (ARIMA)

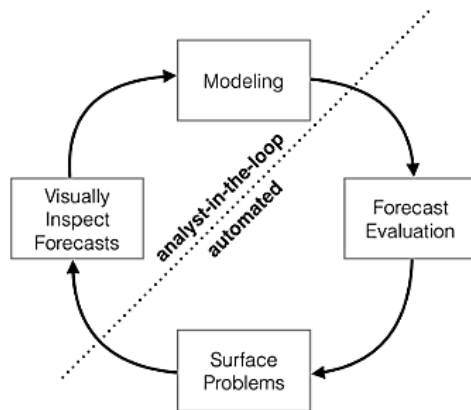
Model ARIMA sama seperti model deret waktu lain, akan sesuai bila peramal dapat mengasumsikan jumlah yang didapat dengan pola masa depan dan tren yang menyerupai pola dan tren saat ini [14]. ARIMA cocok untuk peramalan jangka pendek, 12 bulan atau kurang. ARIMA berbeda dengan *Exponential Smoothing Model* karena *Exponential Smoothing Model* memasukkan jenis tambahan informasi tentang data. Identifikasi model ARIMA didasarkan pada perhitungan *Autocorrelation*, sedangkan model Box-Jenkins lebih ke arah matematika yang kompleks oleh karena itu lebih sulit memahami konseptual.

Model *Non-Seasonal Box-Jenkins* dapat disimbolkan dengan ARIMA(p,d,q) dimana *p* adalah *Autoregressive* (AR), *q* adalah *Moving Average* (MA), dan *d* adalah jumlah data *difference* dengan *de-trend* atau sebaliknya [14]. Dengan *Exponential smoothing*, peramal harus dapat menentukan apakah ada tren dan pola musiman dalam data historis dan kemudian memilih bentuk yang sesuai dari atribut tersebut.

Dengan menggunakan model ARIMA, tugas dari peramal adalah mencari variabel AR, MA, dan *difference* yang sesuai. Secara teori variabel tersebut bernilai *integer* dan pada umumnya bernilai 0, 1, 2 dan 3.

#### F. Facebook Prophet (Prophet)

Dalam paper yang dibahas oleh Benjamin Letham [6], pendekatan peramalan berskala dapat digambarkan seperti pada Gambar 2.



Gambar 2. Skematik pendekatan forecasting at scale

Model yang digunakan oleh *Prophet* menggunakan *decomposable time series model* dengan tiga model komponen, yaitu: *trend*, *seasonal*, dan *irregular components*. Sehingga dapat dirumuskan sebagai berikut:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \quad (4)$$

Dimana setiap komponen persamaan (4), terdiri dari:

- $g(t)$  adalah kurva pertumbuhan piecewise linear atau logistik untuk pemodelan perubahan non-periodik dalam seri waktu
- $s(t)$  adalah perubahan periodik (misalnya musiman mingguan/tahunan)
- $h(t)$  adalah Efek liburan (pengguna disediakan) dengan jadwal yang tidak teratur
- $\varepsilon(t)$  adalah istilah kesalahan account untuk setiap perubahan yang tidak biasa tidak ditampung oleh model

*Prophet* menggunakan waktu sebagai repesor dan mencoba untuk menyesuaikan beberapa persamaan deret waktu linear dan non linear sebagai komponen fungsi waktu. Pemodelan musiman dianggap sebagai komponen aditif dimana pendekatan yang sama diambil oleh *Exponential Smoothing* dalam teknik *Holt-Winters* [15]. *Prophet* meringkaskan masalah peramalan sebagai kurva pas daripada dilihat secara eksplisit pada saat ketergantungan berbasis setiap observasi dalam seri waktu.

#### G. Evaluasi Performa Peramalan

Pada umumnya test data diambil 20% dari total data yang ada [16]. Hal yang perlu diperhatikan adalah:

- Sebuah model yang sesuai dengan data latihan tidak selalu menghasilkan perkiraan dengan baik
- Kecocokan dapat diperoleh dengan menggunakan model yang memiliki parameter yang cukup
- *Overfitting* model menyebabkan data gagal untuk mengidentifikasi pola sistematis

Evaluasi yang digunakan dalam penelitian ini menggunakan:

#### 1) Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2} \quad (5)$$

RMSE dihitung dengan mengkuadratkan error (prediksi – observasi) dibagi dengan jumlah data, lalu diakarkan. Nilai RMSE rendah menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model perkiraan mendekati variasi nilai observasinya. Semakin kecil nilai RMSE, semakin dekat nilai yang diprediksi dan diamati.

2) *Metode AIC dan BIC*: Selain menggunakan tiga hal tersebut, ada metode yang digunakan untuk membuktikan apakah nilai RMSE, MSE, dan MAE memiliki keakuratan yang benar (tidak overfitting atau underfitting), yaitu Akaike Information Criterion (AIC) dan Bayesian Information Criterion (BIC).

Untuk menghitung nilai AIC dapat dirumuskan dengan persamaan sebagai berikut:

$$AIC = e^{\frac{2k}{n} \sum_{i=1}^n \hat{u}_i^2} = \frac{2k}{n} + \ln \left( \frac{\sum_{i=1}^n \hat{u}_i^2}{n} \right) \quad (6)$$

Dimana  $k$  adalah jumlah parameter yang diestimasi dalam model regresi,  $n$  adalah jumlah observasi, dan  $\hat{u}_i$  adalah residualnya. Sedangkan untuk persamaan pada BIC/SIC adalah sebagai berikut:

$$BIC = \frac{k}{n} \frac{\sum_{i=1}^n \hat{u}_i^2}{n} = \left( \frac{k}{n} \right) \ln n + \ln \left( \frac{\sum_{i=1}^n \hat{u}_i^2}{n} \right) \quad (7)$$

Model regresi yang terbaik adalah model yang mempunyai nilai AIC dan BIC terkecil [17].

## IV. HASIL PENELITIAN

### A. Persiapan Data

Berdasarkan dari data yang diterima, data seluruh mahasiswa yang mendaftar ke UK. Maranatha dimana data dimulai dari tahun 1993 sampai 2019 sebanyak 159.770 data calon mahasiswa baru yang dapat disebut juga sebagai *Raw Data* dengan struktur data pada Gambar 3.

| Data Mahasiswa |                     |
|----------------|---------------------|
| PK             | NO_UJIAN            |
|                | TAHUN               |
|                | PIL_UTAMA           |
|                | DESKRIPSI_PIL_UTAMA |
|                | KODE_SKL            |
|                | NAMA_KO_KA          |
|                | PROPINSI            |

Gambar 3. Struktur tabel Data Mahasiswa

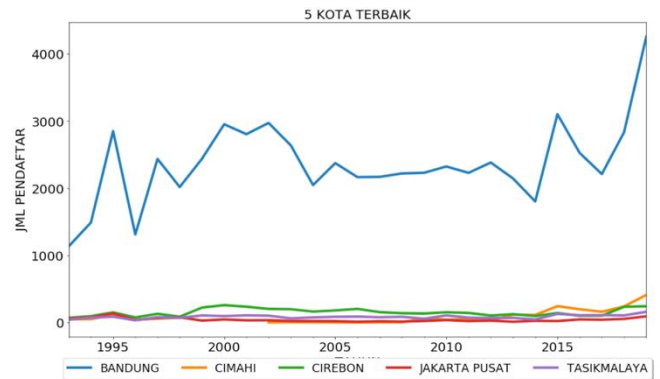
Dari Gambar 3 dipergunakan dua field yaitu TAHUN dan NAMA\_KO\_KA. Dari *Raw Data*, data calon mahasiswa yang mendaftar dilakukan subtotal yang dikelompokkan berdasarkan Tahun dan Nama Kota, sehingga menghasilkan 26 baris data observasi tahunan.

Penelitian ini menggunakan sampel lima kota yang teratas dimana satu kota adalah satu dataset untuk pengujian, sehingga dengan menggunakan asumsi lima sampel data dimungkinkan memenuhi perbandingan pengujian performa *Prophet* dan ARIMA. Hasil dari persiapan data dapat dilihat seperti pada Gambar 4.

| TAHUN | BANDUNG | CIMAHI | CIREBON | JAKARTA PUSAT | TASIKMALAYA |
|-------|---------|--------|---------|---------------|-------------|
| 1993  | 1133.0  | 48.0   | 70.0    | 58.0          | 42.0        |
| 1994  | 1488.0  | 51.0   | 95.0    | 78.0          | 72.0        |
| 1995  | 2849.0  | 109.0  | 151.0   | 136.0         | 87.0        |
| 1996  | 1311.0  | 44.0   | 76.0    | 38.0          | 33.0        |
| 1997  | 2435.0  | 58.0   | 130.0   | 69.0          | 81.0        |
| 1998  | 2017.0  | 73.0   | 87.0    | 77.0          | 69.0        |
| 1999  | 2437.0  | NaN    | 222.0   | 29.0          | 106.0       |
| 2000  | 2951.0  | NaN    | 259.0   | 46.0          | 96.0        |
| 2001  | 2804.0  | NaN    | 235.0   | 33.0          | 107.0       |
| 2002  | 2971.0  | 2.0    | 203.0   | 34.0          | 102.0       |
| 2003  | 2637.0  | 6.0    | 198.0   | 23.0          | 62.0        |
| 2004  | 2046.0  | 4.0    | 163.0   | 22.0          | 76.0        |
| 2005  | 2373.0  | 3.0    | 180.0   | 20.0          | 85.0        |
| 2006  | 2164.0  | 2.0    | 203.0   | 12.0          | 88.0        |
| 2007  | 2168.0  | 4.0    | 155.0   | 19.0          | 78.0        |
| 2008  | 2219.0  | 4.0    | 139.0   | 14.0          | 85.0        |
| 2009  | 2229.0  | 59.0   | 134.0   | 22.0          | 56.0        |
| 2010  | 2323.0  | 42.0   | 153.0   | 37.0          | 106.0       |
| 2011  | 2228.0  | 51.0   | 143.0   | 21.0          | 73.0        |
| 2012  | 2382.0  | 27.0   | 104.0   | 28.0          | 66.0        |
| 2013  | 2145.0  | 117.0  | 124.0   | 11.0          | 69.0        |
| 2014  | 1803.0  | 113.0  | 97.0    | 25.0          | 46.0        |
| 2015  | 3101.0  | 243.0  | 142.0   | 22.0          | 128.0       |
| 2016  | 2525.0  | 198.0  | 100.0   | 46.0          | 108.0       |
| 2017  | 2210.0  | 160.0  | 101.0   | 42.0          | 110.0       |
| 2018  | 2833.0  | 239.0  | 234.0   | 54.0          | 106.0       |
| 2019  | 4255.0  | 411.0  | 242.0   | 92.0          | 160.0       |

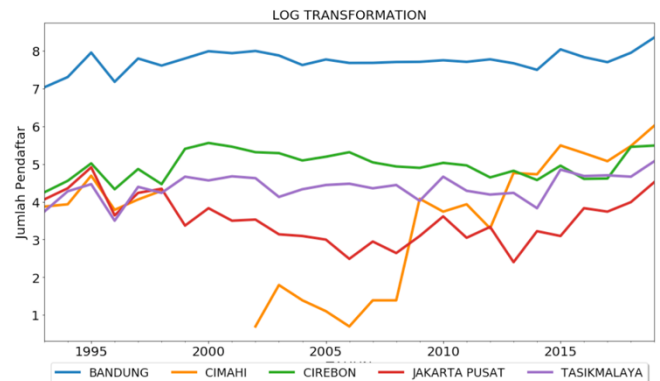
Gambar 4. Data sample lima kota terbaik

Dari Gambar 4, data per kota dapat digambarkan ke dalam grafik sebagai berikut:



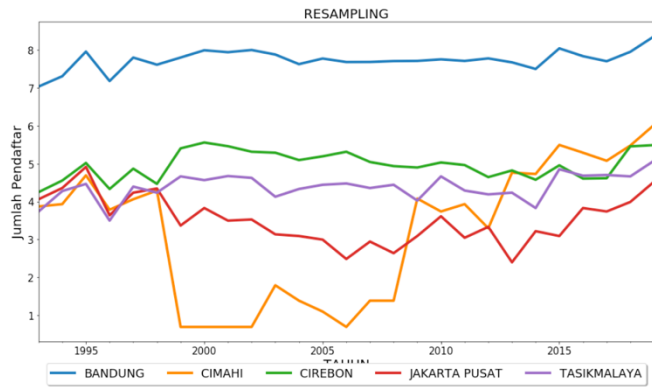
Gambar 5. Grafik dari Sampel Data lima kota terbaik

Terlihat bahwa Bandung memiliki [2] yang berbeda dengan kota lain (Gambar 5), maka akan dilakukan transformasi data dengan menggunakan Logaritma natural maka menghasilkan seperti pada Gambar 6.



Gambar 6. Kondisi setelah dilakukan Log Transformation

Untuk mengatasi ada beberapa tahun yang memiliki nilai kosong, maka dilakukan interpolasi data dengan menggunakan *resampling time-series* dengan fungsi agregasi nilai tengah (*mean*), dimana metode *resampling time-series* [18] digunakan untuk konversi frekuensi deret waktu. Sehingga dengan menggunakan *resampling time-series* maka menghasilkan seperti pada Gambar 7.



Gambar 7. Kondisi setelah dilakukan Resampling Data

### B. ARIMA

Tabel I, digunakan sebagai batasan *range* dalam penelitian ini. Batasan *range* didapatkan dengan cara analisis tentatif dari hasil percobaan transformasi data yang dievaluasi dengan pengujian stasioner *ADF Test* dan *KPSS Test*.

Untuk pencarian model ARIMA(p,d,q) yang memiliki nilai RMSE dan AIC yang terkecil, akan menggunakan pengujian dengan data *training* dan *testing* sebesar 70:30. Pencarian ini berfungsi untuk mencari model ARIMA dengan performa peramalan yang baik.

TABEL I  
RANGE MAKSIMAL MODEL ARIMA

| Kota          | p | d | q |
|---------------|---|---|---|
| Bandung       | 4 | 2 | 5 |
| Cimahi        | 4 | 2 | 5 |
| Cirebon       | 4 | 2 | 5 |
| Jakarta Pusat | 4 | 2 | 5 |
| Tasikmalaya   | 4 | 1 | 5 |

Untuk pengujian performanya, akan dipergunakan RMSE dan divalidasi oleh *Akaike Information Criterion* (AIC), sebab nilai RMSE kecil belum tentu nilai tersebut ideal (ideal dalam arti tidak *overfitting* atau *underfitting*). Sehingga untuk menguji model yang sudah memiliki performa yang baik diperlukan pengujian untuk validasi model tersebut, salah satunya menggunakan metode AIC.

TABEL II  
HASIL EVALUASI MODEL ARIMA

| Kota          | ARIMA |   |   | Hasil Evaluasi |        |
|---------------|-------|---|---|----------------|--------|
|               | p     | d | q | RMSE           | AIC    |
| Bandung       | 4     | 0 | 1 | 0.279          | 2.2632 |
| Cimahi        | 0     | 1 | 0 | 0.643          | 77.107 |
| Cirebon       | 1     | 0 | 0 | 0.331          | 21.740 |
| Jakarta Pusat | 1     | 1 | 0 | 0.492          | 38.493 |
| Tasikmalaya   | 0     | 0 | 4 | 0.461          | 18.089 |

Dari lima data deret waktu (Tabel II), masing-masing kota sudah dioptimasi performa nilai RMSEnya dimana pengujiannya dilakukan dari model (0,0,0) sampai dengan nilai maksimal yang telah ditentukan pada Tabel I.

### C. Prophet

Metode *Prophet* lebih sederhana daripada model ARIMA, karena tidak memerlukan penentuan nilai AR, MA, dan *difference*. Diluar dari komponen utama yang dibutuhkan oleh metode *Prophet*, *Prophet* juga memerlukan pengaturan *changepoint* yang berupa *changepoint prior scale* dan *changepoint range*. *Changepoint* digunakan untuk mendeteksi probabilitas distribusi dari proses *stochastic* atau perubahan dalam deret waktu.

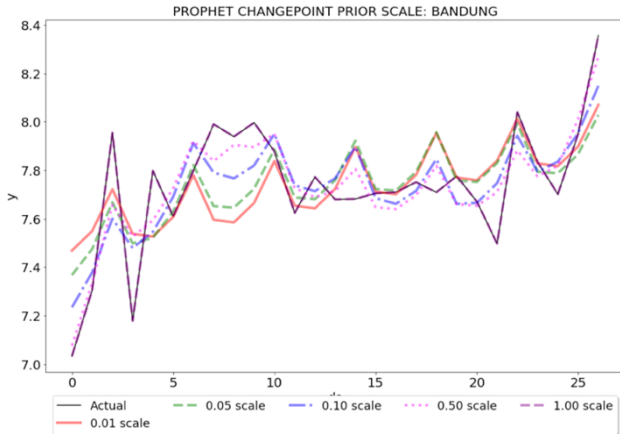
Dalam penelitian ini, untuk menentukan *changepoint prior scale* menggunakan asumsi subjektif dimana menggunakan 5 jenis *changepoint prior scale*, yaitu 0.01, 0.05, 0.10, 0.50, dan 1. Sedangkan untuk *changepoint range*, dibatasi dengan menggunakan nilai 1, karena data yang digunakan hanya 26.

Ada tiga kondisi dalam *Prophet*, yaitu kondisi *underfitting*, normal, dan *overfitting*. Dengan menggunakan lima jenis *changepoint prior scale*, performa RMSE akan dibandingkan (lihat Tabel III), untuk mendapatkan model yang tidak *overfitting* atau *underfitting* diperlukan visualisasi (lihat Gambar 8). Dalam penelitian ini, penentuan *changepoint prior scale* dilakukan secara subjektif.

TABEL III  
PROPHET TERHADAP KOTA BANDUNG

| Changepoint Prior Scale | Changepoint Range | RMSE  | AIC       |
|-------------------------|-------------------|-------|-----------|
| 0.010                   | 1                 | 0.215 | -4.363    |
| 0.050                   | 1                 | 0.198 | -8.781    |
| 0.100                   | 1                 | 0.160 | -20.319   |
| 0.500                   | 1                 | 0.135 | -29.535   |
| 1.000                   | 1                 | 0.000 | -1633.928 |

Untuk menentukan model mana yang akan digunakan dalam peramalan, maka dapat ditentukan dengan melihat nilai RMSE-nya, semakin kecil nilai RMSE ada kemungkinan *overfitting*, sebaliknya semakin besar nilai RMSE ada kemungkinan *underfitting*. Sedangkan untuk nilai AIC yang dihasilkan semakin besar *Changepoint Prior Scale*, maka nilai AIC akan semakin kecil.



Gambar 8. *Changepoint Prior Scale* Kota Bandung

Berdasarkan asumsi tentatif yang dilakukan terhadap grafik dari perbandingan setiap jenis *changepoint prior scale*, model *prophet* (Tabel IV) akan digunakan untuk dilakukan pengujian terhadap data ramalan.

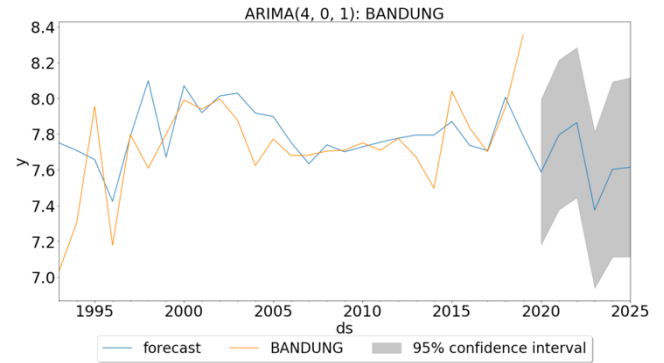
TABEL IV  
HASIL EVALUASI MODEL *PROPHET*

| Kota          | CP Scale | CP Range | Hasil Evaluasi |         |
|---------------|----------|----------|----------------|---------|
|               |          |          | RMSE           | AIC     |
| Bandung       | 0.500    | 1        | 0.135          | -29.535 |
| Cimahi        | 1.000    | 1        | 0.688          | 58.408  |
| Cirebon       | 1.000    | 1        | 0.122          | -34.878 |
| Jakarta Pusat | 1.000    | 1        | 0.291          | 11.965  |
| Tasikmalaya   | 1.000    | 1        | 0.232          | 0.272   |

Jika dibandingkan ARIMA (Tabel II) dengan *Prophet* (Tabel IV), 4 kota dengan metode *Prophet*, lebih unggul dari ARIMA, sedangkan pada Cimahi, ARIMA lebih unggul dari *Prophet*. Untuk membuktikan apakah kedua metode itu tidak mengalami *overfitting* atau *underfitting*, dilakukan analisis visual yang terbentuk dari hasil ramalan dengan model yang sudah ditentukan, yaitu untuk visualisasi hasil peramalan ARIMA mempergunakan model Tabel II dan visualisasi hasil peramalan *Prophet* mempergunakan model Tabel IV.

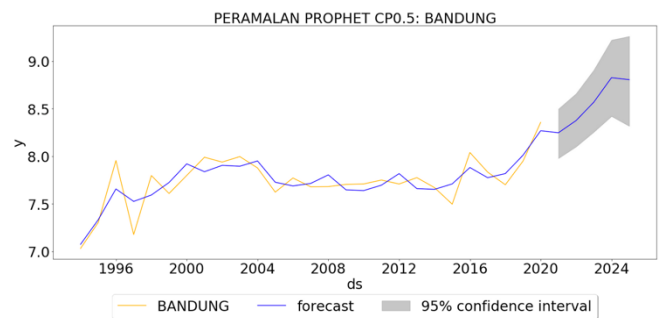
#### D. Forecasting

Untuk Kota Bandung (Gambar 9), hasil ramalan yang dihasilkan oleh model ARIMA(4,0,1) dengan tingkat performansi RMSE sebesar 0.2792.



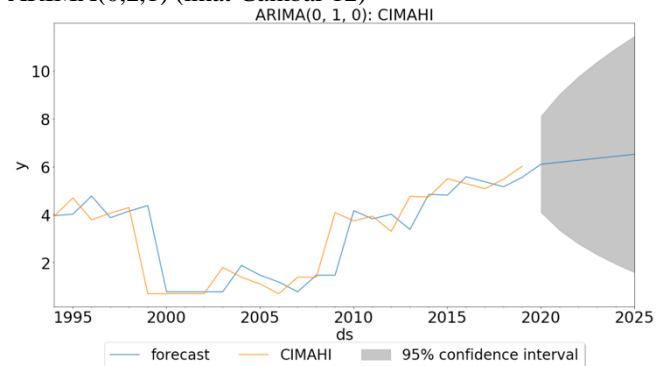
Gambar 9. Forecast Kota Bandung dengan ARIMA

Pada grafik tersebut dapat dilihat bahwa ada lonjakan cukup signifikan pada tahun 2017-2019, sehingga *Confidence Interval* (batas dari ramalan yang akan terjadi) belum menyesuaikan dari lonjakan tersebut. Lain halnya apabila menggunakan model *Prophet* dengan *changepoint prior scale* 0.5 (Gambar 10), metode *prophet* menghasilkan performa RMSE sebesar 0.135. Dan dapat dilihat dari grafik tersebut, bahwa ramalan *Prophet* dapat beradaptasi dengan data aktual meskipun ada lonjakan pada tahun 2017-2019.



Gambar 10. Forecast Kota Bandung *Prophet* cp0.5

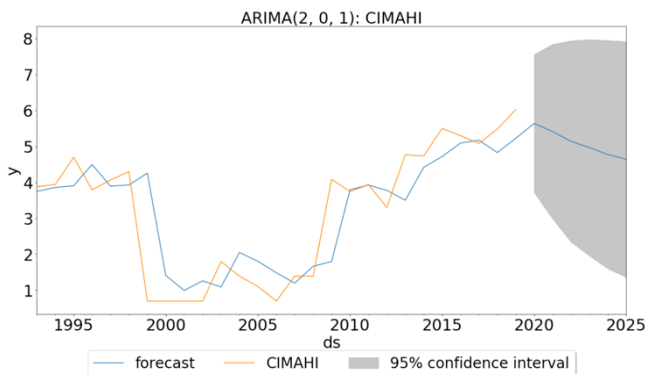
Pada Cimahi, apabila menggunakan model ARIMA (0,1,0) (lihat Gambar 11) menunjukkan pola *overfitting* dimana *confidence interval* membesar, meskipun memiliki RMSE dan AIC terkecil dari model lainnya (RMSE 0.6434 dan AIC 77.1066), sehingga untuk menghasilkan pola yang tidak *overfitting* maka mempergunakan model ARIMA(0,2,1) (lihat Gambar 12)



Gambar 11. Forecast Cimahi model (0,1,0)

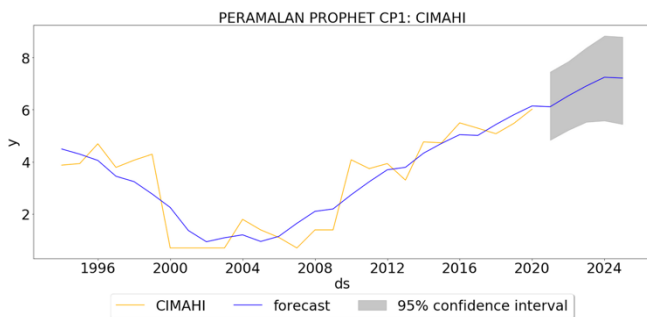


Model (0,2,1) sudah tidak mengalami *overfitting*, sehingga model ini dapat digunakan sebagai model peramalan untuk Cimahi dengan resiko penurunan performa RMSE sebesar 0.044.



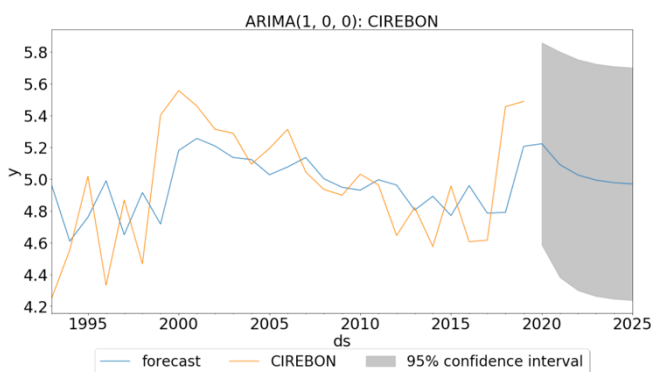
Gambar 12. Forecast Cimahi model (0,2,1)

Berbeda dari *Prophet* (Gambar 13), dengan menggunakan *changepoint prior scale* 1, tidak mengalami *overfitting* dan menghasilkan performa RMSE sebesar 0.688.



Gambar 13. Forecast Cimahi Prophet cp1

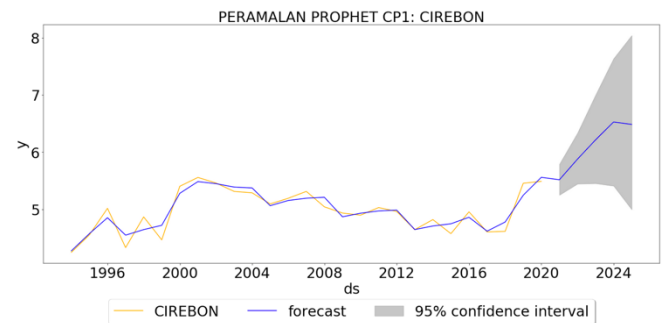
Untuk Cirebon, ARIMA tidak mengalami pola *overfitting* seperti kota Cimahi (Gambar 14), akan tetapi untuk *Prophet*, Cimahi mengalami *overfitting* (Gambar 15).



Gambar 14. Forecast Cirebon model (1,0,0)

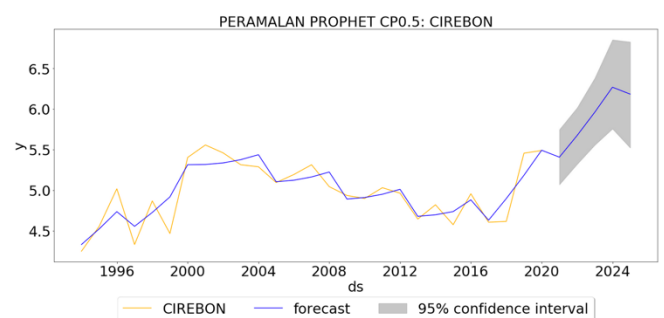
Hasil ramalan yang dihasilkan oleh model ARIMA(1,0,0) dengan tingkat performansi RMSE sebesar 0.331. Pada grafik itu tampak normal meskipun menggunakan model

ARIMA yang hanya menggunakan AR. Jika melihat *Confidence Interval* (batas dari ramalan yang akan terjadi) yang tampak besar pada ramalan, itu disebabkan karena jika dilihat antara tahun 1993 – 2019 memiliki peningkatan dan penurunan yang tinggi.



Gambar 15. Forecast Cirebon Prophet cp1

Metode *Prophet* menghasilkan performa RMSE sebesar 0.122. Akan tetapi jika dilihat dari *Confidence Interval* (batas dari ramalan yang akan terjadi) mengalami pembesaran pada tahun 2023. Sehingga ada kemungkinan *changepoint prior scale* 1 mengalami *overfitting*. Untuk mengatasi hal ini dapat mempergunakan *changepoint prior scale* lainnya yaitu *changepoint prior scale* 0.5 (Gambar 16).

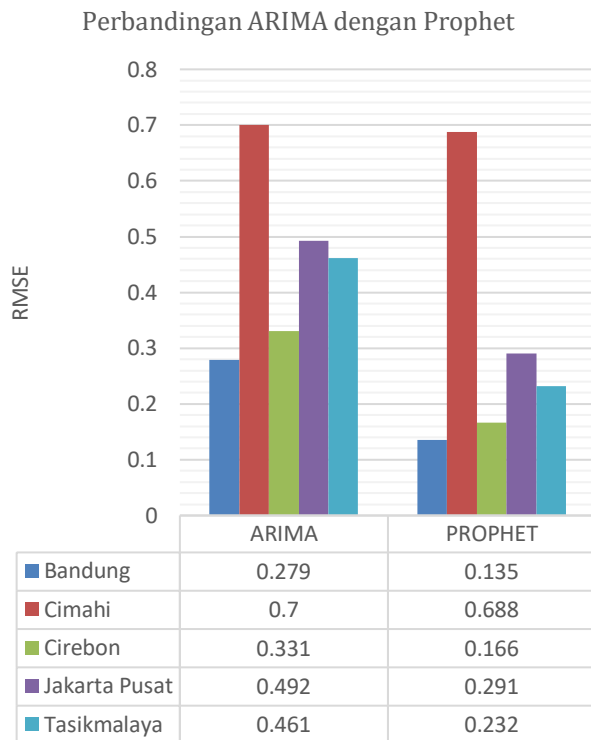


Gambar 16. Forecast Cirebon Prophet cp0.5

Metode *Prophet* dengan *changepoint prior scale* 0.5 tidak mengalami *overfitting*, akan tetapi performa RMSE menjadi 0.166, menurun sebesar 0.044 point dari *changepoint prior scale* 1.

Dari percobaan dengan hasil peramalan, maka dapat disimpulkan bahwa RMSE dan AIC terkecil belum tentu menghasilkan ramalan yang ideal (tidak *overfitting* atau *underfitting*).

Sehingga perbandingan akhir untuk model ARIMA dan *Prophet* (Gambar 17) menyatakan bahwa *Prophet* mampu berperforma lebih baik daripada ARIMA.



Gambar 17 Perbandingan ARIMA dengan Prophet

## V. KESIMPULAN DAN SARAN

### A. Kesimpulan

Kedua model dilatih dengan menggunakan 26 data dan 5 sampel data deret waktu. Dari hasil tersebut, menghasilkan kesimpulan bahwa Model *Prophet* mampu berkinerja lebih unggul dari ARIMA berdasarkan dari nilai RMSE (Gambar 17), sehingga hipotesis  $H_1$  dapat diterima dan untuk peramalan terhadap data pemasaran mahasiswa UK. Maranatha, dapat mempergunakan metode *Prophet* sebagai metode yang ditawarkan untuk membantu tim pemasaran dalam melakukan peramalan jumlah pendaftar mahasiswa baru per kota, dimana dari hasil ramalan tersebut dapat digunakan dalam pengambilan keputusan nantinya.

### B. Saran

Dengan adanya beberapa kekurangan yang dilakukan dalam penelitian ini, dapat disimpulkan beberapa saran yang bertujuan untuk meningkatkan penelitian selanjutnya, salah satunya adalah memperhatikan dataset yang digunakan, apabila dataset yang digunakan hanya menggunakan variabel per tahun (*annual*) sebagai data observasi deret waktu, akan menyebabkan observasi data untuk *Training* dan *Testing* model data menjadi kurang maksimal. Sedangkan penentuan model, tidak dapat dilihat dari RMSE dan AIC terkecil saja, melainkan harus di implementasikan

ke dalam hasil ramalan dan dilakukan visualisasi dari model ramalan tersebut agar tidak terjadi *overfitting* ataupun *underfitting*. Untuk metode *Prophet*, penentuan *changepoint prior scale* dan *changepoint range* diperlukan, karena apabila *changepoint prior scale* dan *range* tidak diatur, maka metode *Prophet* tidak dapat menghasilkan nilai yang optimum (akan mengalami *underfitting* atau *overfitting*).

## DAFTAR PUSTAKA

- [1] R. Ginting, *Sistem Produksi*, Graha Ilmu, Yogyakarta, 2007.
- [2] P. Landau. (2018) Business Forecasting: Why You Need It & How to Do It. [Online]. Tersedia: <https://www.projectmanager.com/blog/business-forecasting>
- [3] C. Polat, "Forecasting as a Strategic Decision-Making Tool: A Review and Discussion with Emphasis on Marketing Management," *Euro Journals Publishing Inc.*, vol. 20, no. 2, pp. 423-425, 2008.
- [4] NIST. Introduction to Time Series Analysis. [Online]. Tersedia: <https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm>
- [5] A. H. M. Imrul Hasan & P. Haddawy, "Integrating ARIMA and Spatiotemporal Bayesian Networks for High Resolution Malaria Prediction," in *2016 ECAI 22nd European Conference on Artificial Intelligence*, The Hague, Netherlands, IOS Press, 2016, pp. 1783-1790, doi: 10.3233/978-1-61499-672-9-1783.
- [6] S. J. Taylor & B. Letham, "Forecasting at Scale," *PeerJ Preprints*, 2017, doi: 10.7287/peerj.preprints.3190v2.
- [7] I. Yenidogan, A. Cayir, O. Kozan, T. Dag, & C. Arslan, "Bitcoin Forecasting using ARIMA and Prophet," in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, Sarajevo, IEEE, 2018, pp. 621-624, doi: 10.1109/UBMK.2018.8566476.
- [8] R. J. Hyndman & A. V. Kostenko, "Minimum Sample Size Requirements for Seasonal Forecasting Models," *Foresight: The International Journal of Applied Forecasting*, International Institute of Forecasters, no. 6, pp. 12-15, Spring 2007.
- [9] W. Stevenson, *Operation Management*, 11th ed., McGraw-Hill, New York, 2012, pp. 74-77.
- [10] J. C. Chambers, S. K. Mullick, & D. D. Smith. (1971) How to Choose the Right Forecasting Technique. [Online]. Tersedia: <https://hbr.org/1971/07/how-to-choose-the-right-forecasting-technique>
- [11] A. Dickey & W. A. Fuller, "Distribution of the Estimator for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 427-431, 1979, doi: 10.2307/2286348.
- [12] W. Wang, *Stochasticity, Nonlinearity, and Forecasting of Streamflow Processes*, Delft University Press, 2006, pp. 34-36.
- [13] Kwiatkowski, P. C. Phillips, P. Schmidt, & Y. Shin, "Testing the null hypothesis of stationary against the alternative of a unit root," *Journal of Econometric*, vol.54, no. 1-3, pp. 159-178, 1992, doi: 10.1016/0304-4076(92)90104-Y
- [14] Stellwagen & L. Tashman, "ARIMA: The models of Box and Jenkins," *Foresight: The International Journal of Applied Forecasting*, no. 30, pp. 28-33, 2013.
- [15] Solarwinds. (2019) Holt-Winters Forecasting Simplified. [Online], Tersedia: <https://orangematter.solarwinds.com/2019/12/15/hot-winters-forecasting-simplified/>
- [16] R. J. Hyndman & G. Athanasopoulos, *Forecasting: Principles and Practice*, OText, Melbourne, 2018, pp. 164-165.
- [17] A. Widarjono, *Ekonomika: Teori dan Aplikasi untuk Ekonomi dan Bisnis*, 2nd ed., Ekonisia: Fakultas Ekonomi Universitas Islam Indonesia, Yogyakarta, 2007.
- [18] Pandas Development Team, (2014) Pandas API Documentation. [Online] Tersedia: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.resample.html>