

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Deret waktu (*time series*) merupakan nilai observasi yang berurutan pada suatu interval waktu tertentu berdasarkan satuan waktu. Analisis deret waktu memiliki berbagai manfaat yang aplikatif dalam kehidupan sehari-hari, yaitu pada aspek peramalan atau *forecasting*, sehingga dapat dijadikan sebagai dasar perencanaan masa depan. Peramalan merupakan proses perkiraan nilai di masa depan berdasarkan informasi yang diketahui, yaitu data deret waktu pada masa lampau [1].

Kelemahan model deret waktu dalam melakukan peramalan umumnya datang dari faktor-faktor seperti adanya *trend* dan *seasonalitas*. Dalam proses analisis deret waktu, kedua masalah ini dapat diatasi dengan melakukan *differencing* dan transformasi supaya deret waktu yang dimiliki menjadi deret waktu yang memenuhi asumsi stasioneritas. Deret waktu dapat dikatakan sebagai deret waktu stasioner apabila sebaran datanya berada di sekitar nilai rata-rata yang konstan, dan memiliki variansi yang sama pada seluruh periode waktu. Apabila deret waktu telah stasioner, maka proses peramalan baru dapat dilakukan pada deret waktu tersebut.

Metode analisis deret waktu yang umum digunakan adalah metode *Box-Jenkins* yang dikemukakan oleh George E. P. Box dan Gwilym M. Jenkins pada tahun 70-an. Kedua matematikawan ini memperkenalkan sebuah model bernama ARIMA (*Auto Regressive Integrated Moving Average*) yang diimplementasikan dengan cara melakukan *fitting* data deret waktu terhadap model ARMA (*Auto Regressive Moving Average*) jika stasioner, dan model ARIMA jika tidak stasioner. Model umum ARMA adalah model gabungan AR (*Auto Regressive*) dan MA (*Moving Average*). Model AR terdiri dari variabel tidak bebas yang dipengaruhi oleh dirinya sendiri pada indeks waktu lampau, sedangkan model MA merupakan regresi linier dari nilai saat ini dengan suku galat saat ini dan indeks waktu sebelumnya. Komponen I (*Integrated*) dalam model 'ARIMA' merupakan berapa kali dilakukannya proses *differencing* terhadap data awal sebelum data tersebut menjadi data deret waktu yang memenuhi asumsi stasioneritas.

Model ARIMA umumnya masih mengandung suku *error* berupa *white noise*. *Error* merupakan suku yang nilainya ingin diminimalisir dalam sebuah

model. Berdasarkan penelitian terdahulu [2] dan [3], dikatakan bahwa penerapan algoritma *Kalman filter* dapat memperbaiki model ARIMA sehingga meminimalisir *error*.

Algoritma *Kalman filter* mendeskripsikan solusi rekursif untuk masalah pemfilteran linier dari data diskrit [4]. Dalam prosesnya, model ditransformasikan ke dalam representasi *state space* dengan *input* orde dan parameter model deret waktu. Setelah itu, dilakukan tahap *prediction-update* secara rekursif untuk mendapatkan estimasi parameter model yang telah terfilter.

Oleh karena itu, dalam penelitian ini dilakukan peramalan menggunakan dua model, model deret waktu ARIMA dan model ARIMA-KF (yaitu model ARIMA yang dikenakan algoritma *Kalman filter*). Kedua model dibandingkan dengan indikator akurasi RMSE, MAE, dan MAPE. Dilanjutkan dengan melakukan regresi logistik [5] untuk memprediksi suatu variabel dependen menggunakan beberapa variabel independen dari hasil peramalan model deret waktu.

Sebagai studi kasus, data suhu dan kelembaban sebagai unsur cuaca (suhu dan kelembaban merupakan dua faktor utama yang berperan besar dalam menentukan cuaca yang terjadi [6]), serta data curah hujan sebagai cuaca di Bandar Lampung digunakan menjadi sampel data penelitian. Alasan terpilihnya Bandar Lampung adalah karena struktur penduduk bekerja menurut lapangan pekerjaan pada Februari 2019 masih didominasi oleh tiga lapangan pekerjaan utama, yaitu: Pertanian, Kehutanan, dan Perikanan sebesar 47,97% [7], dimana berlangsungnya kegiatan tersebut dipengaruhi oleh cuaca yang terjadi. Selain itu, Lampung menduduki peringkat ketiga provinsi di Pulau Sumatera dengan rumah tangga usaha pertanian terbesar pada 2013 [8].

## 1.2 Rumusan Masalah

Penulisan skripsi ini menjawab rumusan masalah sebagai berikut.

1. Bagaimana hasil perbandingan peramalan unsur-unsur cuaca dengan model ARIMA dan ARIMA-KF?
2. Bagaimana regresi logistik dapat memprediksi cuaca dengan menggunakan variabel peramalan unsur-unsur cuaca?
3. Bagaimana perbandingan hasil prediksi cuaca antar model regresi logistik dan perbandingan hasil prediksi cuaca dengan data aktual cuaca?

### 1.3 Tujuan Penelitian

Penulisan skripsi ini memiliki tujuan sebagai berikut.

1. Membandingkan hasil peramalan unsur-unsur cuaca dari model ARIMA dan ARIMA-KF dengan menggunakan tiga indikator akurasi.
2. Memprediksi cuaca berdasarkan hasil peramalan unsur-unsur cuaca dengan menggunakan regresi logistik biasa dan polinomial.
3. Membandingkan hasil prediksi cuaca antar model regresi logistik biasa dan polinomial serta membandingkan hasil prediksi cuaca dengan data aktual cuaca.

### 1.4 Batasan Masalah

Batasan masalah dan asumsi pada skripsi ini adalah sebagai berikut.

1. Data suhu, kelembaban udara, dan curah hujan yang digunakan adalah data di Bandar Lampung dari Januari 2015 sampai September 2020.
2. Unsur-unsur cuaca yang diperhitungkan dalam skripsi ini hanya suhu dan kelembaban, tidak mencakup unsur-unsur cuaca lainnya.
3. *Missing values* pada data suhu dan kelembaban diasumsikan merupakan rata-rata suhu dan kelembaban pada tanggal yang sama dari rentang tahun data yang digunakan.
4. Cuaca yang diprediksi dalam skripsi ini terdiri dari hujan dan cerah (tidak hujan), tidak termasuk berawan, panas, sejuk, maupun berangin.
5. Data cuaca didapatkan dari klasifikasi biner data curah hujan. Ada atau tidaknya curah hujan berkorespondensi dengan cuaca hujan (1) atau cerah/tidak hujan (0).
6. *Missing values* pada data curah hujan tidak digunakan, sehingga dilakukan pembersihan untuk setiap baris data (termasuk suhu dan kelembaban di tanggal tersebut) yang mengandung *missing values*.
7. Orde ( $p$  dan  $q$ ) maksimum pada model deret waktu adalah 5.
8. Tiga bulan terpilih untuk diprediksi cuacanya adalah bulan Januari, Februari, dan September 2020.

9. *Threshold* yang digunakan dalam regresi logistik adalah 0.6.
10. Taraf signifikansi ( $\alpha$ ) yang digunakan dalam uji hipotesis diasumsikan sebesar 5%.
11. Penggunaan kata "peramalan" dalam skripsi ini merujuk pada hasil dari model deret waktu sedangkan kata "prediksi" merujuk pada hasil dari model regresi logistik.
12. Akumulasi *error* dari peramalan model deret waktu tidak diperhitungkan untuk mengukur akurasi prediksi cuaca.

## 1.5 Manfaat

Penulisan skripsi ini memiliki manfaat teoritis dan praktis sebagai berikut.

### 1.5.1 Manfaat Teoritis

Manfaat teoritis dari skripsi ini adalah untuk memahami penggunaan algoritma *Kalman filter* sebagai alat bantu untuk memperbaiki peramalan model deret waktu dan membuktikan bahwa prediksi dapat dilakukan dengan menggunakan data peramalan.

### 1.5.2 Manfaat Praktis

Manfaat praktis dari skripsi ini adalah untuk memprediksi cuaca yang akan terjadi di masa depan agar kegiatan masyarakat di Bandar Lampung dapat berjalan dengan baik. Tidak diperlukannya biaya pengukuran suhu dan kelembaban di masa depan akan menghemat biaya yang dikeluarkan pemerintah untuk memprediksi cuaca.

## 1.6 Sistematika Penulisan

Penulisan skripsi ini memiliki sistematika sebagai berikut.

### 1. BAB I Pendahuluan

Pada bab ini penulis menjelaskan latar belakang pemilihan topik, tujuan yang ingin dicapai, rumusan masalah, batasan-batasan masalah dan asumsi-asumsi yang digunakan, serta manfaat teoritis dan manfaat praktis yang diperoleh

dari skripsi ini akan dipaparkan pada bab ini. Selain itu, pada bab ini juga terdapat penguraian sistematika penulisan skripsi ini.

## 2. **BAB II Landasan Teori**

Pada bab ini penulis menjelaskan teori-teori dasar yang digunakan sebagai landasan skripsi ini. Teori-teori yang akan dijelaskan adalah teori mengenai deret waktu, stasioneritas, pembentukan model deret waktu, model deret waktu stasioner, peramalan deret waktu, *Kalman filter*, *k-folds cross validation*, regresi logistik, dan berbagai indikator akurasi peramalan. Kemudian, penulis memaparkan secara singkat studi-studi yang sudah pernah dilakukan di masa lalu sebagai referensi dalam pengembangan model untuk menjawab tujuan skripsi akan dijelaskan dalam bab ini.

## 3. **BAB III Metode Penelitian**

Pada bab ini penulis memaparkan langkah-langkah yang harus dilakukan untuk menjawab tujuan penulisan skripsi. Tujuan dari skripsi ini adalah melakukan prediksi cuaca cerah atau hujan dengan menggunakan unsur-unsur cuaca, suhu dan kelembaban yang sebelumnya sudah di ramal dengan model deret waktu yang telah diperbaiki dengan *Kalman filter*.

## 4. **BAB IV Analisis dan Pembahasan**

Pada bab ini penulis memaparkan penjelasan mengenai data yang dibutuhkan dalam skripsi ini akan dipaparkan pada bab ini. Kemudian, model ARIMA yang terpilih dan telah diperbaiki oleh *Kalman filter*, disertai orde dan parameternya akan ditunjukkan pada bab ini. Proses peramalan dan hasil indikator akurasi MAPE, MAE, dan RMSE juga akan dipaparkan. Pada akhir bab, analisis terhadap hasil prediksi cuaca akan dilakukan untuk melihat keakuratan hasil prediksi jika dibandingkan dengan cuaca yang terjadi sebenarnya pada hari itu.

## 5. **BAB V Kesimpulan dan Saran**

Pada bab ini, penulis menyimpulkan keseluruhan skripsi ini serta menjawab tujuan-tujuan yang dikemukakan pada Bab I dengan penjelasan yang singkat. Kemudian, penulis menuliskan saran-saran teoritis dan praktis yang bermanfaat bagi pengembangan topik penelitian ini di masa depan.