

Analisis Peramalan dan Pengelompokan Jumlah Turis ke Jepang

Forecasting Analysis and Grouping of the Number of Tourists to Japan

Laurence^{1*}, Devanny Gumulya², J. Sandra Sembel³, Magdalena Lestari Ginting⁴

¹Program Studi Teknik Industri, Universitas Pelita Harapan

²Program Studi Desain Grafis, Universitas Pelita Harapan

³Program Studi Manajemen, Universitas Pelita Harapan

⁴Program Studi Ilmu Komunikasi, Universitas Pelita Harapan

*Penulis korespondensi: Laurence, laurence@uph.edu

Abstrak

Pariwisata merupakan salah satu kontributor penting dalam menunjang perekonomian suatu negara. Penelitian ini menitikberatkan pada kajian kunjungan wisatawan asing ke Jepang dengan mengambil data jumlah wisatawan yang berkunjung dan jumlah pengeluaran wisatawan untuk kategori akomodasi, hiburan, makanan dan minuman, belanja, transportasi, dan lain-lain. Pada studi yang dilakukan sebelumnya tidak terdapat pengelompokan negara untuk berbagai macam pengeluaran ini, sehingga posisi penelitian ini adalah mengisi kekosongan tersebut dengan melakukan pengelompokan negara berdasarkan pengeluaran turis. Selain itu, tujuan studi ini juga membuat model peramalan dengan menggunakan metode ARIMA yang mengakomodasi tren dan musim. Data yang terdiri dari enam jenis pengeluaran direduksi menjadi 2 dengan nilai variansi yang dijelaskan sebesar 83,84%. Hasil pengolahan data menunjukkan 2 kelompok negara turis berdasarkan pengeluarannya. Dua grup tersebut terdiri dari 8 negara anggota OECD dan 12 negara non OECD. Turis yang berasal dari negara yang tergabung dalam OECD memberi memainkan peranan penting dalam perekonomian dunia dengan kontribusi sebesar 50,5 % dari total pengeluaran turis dunia. Kualitas gugus dikategorikan baik dengan rata-rata koefisien siluet dan nilai kohesi 0,56. Pengelompokan ini dapat digunakan sebagai dasar untuk melakukan studi perilaku konsumen setiap negara. Metode peramalan menggunakan ARIMA dapat digunakan dengan memasukan elemen tren dan musim ke dalam model. Nilai R^2 pada model peramalan menunjukan hasil yang baik pada sebagian besar data turis dari 20 negara. Model ARIMA musiman ini dapat dipertimbangkan sebagai model untuk meramalkan jumlah turis yang datang.

Kata kunci: ARIMA, k-means clustering, nilai silhouette coefficient and cohesion, Principal component analysis

Abstract

Tourism is one of the important contributors to supporting a country's economy. This study focuses on studies of foreign tourist visits to Japan by taking data on the number of tourists visiting and the amount of tourist spending for the categories of accommodation, entertainment, food and drink, shopping, transportation, and others. In previous studies, there was no clustering of countries for these various expenditures, so the position of this research is to fill the gap by grouping countries based on tourist expenditures. Apart from that, the purpose of this study is also to create a forecasting model using the ARIMA method that accommodates trends and seasons. The dimensions of the six variables are reduced to 2 with an explained variance of 83.84%. The number of clusters obtained is 2 which divides 20 countries into 8 OECD member countries and 12 countries that are not OECD members. Tourists from countries that are members of the OECD play an important role in the world economy with a contribution of 50.5% of the total world tourist expenditure. The quality of the cluster is categorized as good with the average silhouette coefficient and cohesion value of 0.56. This clustering can be used as a basis for conducting consumer behavior studies from various countries. The forecasting

method using ARIMA can be used by entering trend and season elements into the model. The R^2 value in the forecasting model shows good results in most tourist data from 20 countries. This seasonal ARIMA model can be considered as a model to predict the number of tourists who come.

Keywords: ARIMA, k-means clustering, Principal component analysis, the score of silhouette coefficient and cohesion,

Informasi Artikel:

Diterima tanggal 1 Desember 2020; Disetujui tanggal 4 Desember 2021; Terbit online tanggal 30 Desember 2021

1. Pendahuluan

Pariwisata merupakan salah satu dari industri jasa yang berkembang dengan cepat yang memiliki dampak terhadap berbagai sektor. Laporan dampak ekonomi yang diterbitkan oleh *World Travel & Tourism Council (WTTC)* menunjukkan jumlah yang signifikan dari pariwisata terhadap pemasukan suatu negara dan lowongan kerja. Kontribusi sektor ini yaitu sebesar 10,64% dari total GDP dunia dan menyumbang 1 dari 4 pekerjaan baru yang diciptakan. Oleh karena itu, industri ini sangat menarik untuk dikaji bukan hanya karena besarnya konsumsi dari jumlah wisatawan, penumpang, dan pengunjung tetapi juga karena besarnya pengaruh pembangunan ekonomi terhadap penduduk yang tinggal di tujuan wisata. Secara makro, data prediksi jumlah kedatangan turis dapat digunakan untuk mengestimasi dampak pariwisata pada pertumbuhan ekonomi, dan secara mikro data prediksi dapat digunakan untuk perencanaan operasional terkait pembangunan infrastruktur dan investasi yang dibutuhkan (Frechtling, 2001). Oleh karena itu, data prediksi kedatangan jumlah turis menjadi hal yang penting untuk dilakukan suatu negara. Faktor yang berperan penting untuk mendorong jumlah kunjungan wisatawan yaitu harga tiket pesawat.

Keberadaan maskapai berbiaya murah dapat berkontribusi mendorong wisatawan melakukan kunjungan wisata. Pemilihan menggunakan pesawat seperti ini memiliki dampak yang signifikan terhadap pengambilan keputusan yang berhubungan dengan pengeluaran turis, seperti akomodasi (Silva, Câmara, Vieira, & Santos, 2020). Namun biaya transportasi pesawat saja tidaklah cukup, karena pelancong datang ke suatu daerah karena memiliki ketertarikan. Karena itu keunikan produk lokal yang dimiliki oleh suatu daerah mampu menarik minat wisatawan untuk datang dan berbelanja di sana, seperti dalam studi mengenai industri minuman anggur yang dilakukan oleh (Tafel & Szolnoki, 2020).

Kekhasan lain yaitu panorama yang dimiliki suatu negara yang memiliki empat musim menjadi salah satu destinasi favorit para wisatawan. Data menunjukkan pada tahun 2011 hingga 2015, kedatangan turis asing ke Jepang meningkat 33% per tahun, peningkatan ini merupakan salah satu pertumbuhan tertinggi di dunia (Andonian, Kuwabara, Yamakawa, & Ishida, 2016). Pertumbuhan tersebut didorong oleh beberapa faktor eksternal seperti pertumbuhan ekonomi Tiongkok (30% turis asing yang datang ke Jepang berasal dari Tiongkok) dan peraturan Kementerian Luar Negeri Jepang mengenai pembebasan visa bagi turis asing dari 68 negara dengan paspor elektronik (*e-pasport*), dimana Indonesia dan negara ASEAN seperti Malaysia, Singapura, dan Thailand termasuk dalam 68 negara tersebut.

Pada tahun 2011 terjadi penurunan jumlah wisatawan terjadi akibat bencana gempa dan tsunami di kawasan Tohoku dan peristiwa nuklir di Fukushima. Dengan penurunan tersebut, pemerintah Jepang berusaha untuk mempromosikan pariwisata Jepang dan pada tahun berikutnya jumlah wisatawan meningkat. Umumnya kegiatan promosi yang diadakan oleh sebuah badan pariwisata tidak dilakukan di semua kota di tiap negara, melainkan di tempat-tempat yang berpotensi menarik wisatawan. Hal ini seharusnya bertujuan agar kegiatan

promosi tepat sasaran sehingga lebih efektif dan efisien dalam hal sumber daya yang harus dikeluarkan oleh tim promosi.

Hal tersebut telah dilakukan dengan berbagai kegiatan yang didukung oleh berbagai badan pariwisata di setiap negara. Kegiatan promosi pariwisata di Indonesia berada di bawah koordinasi Kementerian Pariwisata dan Ekonomi Kreatif. Untuk negara lain, contoh badan pariwisata seperti *Japan National Tourism Board*, *Korea Tourism Organization*, *Singapore Tourism Board* juga aktif menyelenggarakan promosi di kota-kota besar di Indonesia. Organisasi-organisasi ini juga memfokuskan pada pentingnya pemasaran bagi wisatawan yang akan berdampak besar terhadap perekonomian, dengan melihat berbagai variabel pendukung seperti misalnya biaya yang dikeluarkan oleh wisatawan per hari (Dwyer & Forsyth, 2008). Biaya tersebut mencakup berbagai aktifitas wisatawan baik untuk memenuhi kebutuhan esensial seperti makanan ataupun keinginan kegiatan sekunder seperti hiburan.

Berbagai jenis kegiatan wisata yang dapat dilakukan oleh wisatawan seperti kegiatan hiburan yaitu mengunjungi tempat-tempat yang belum pernah dikunjungi untuk menambah pengetahuan, juga mengenal lebih jauh tentang budaya, atau juga menjauhi kegiatan rutin dan mencoba gaya hidup dan pengalaman kuliner yang baru (Jang, Morrison, & O'Leary, 2002). Jika mereka melakukan berbagai aktifitas tersebut, pada saat pelancong pulang ke negara asal, mereka dapat berbagi pengalaman perjalanan dengan teman atau keluarga. Selain itu mereka juga bisa memberikan buah tangan yang mereka beli di tempat wisata. Meski wisatawan umumnya berbelanja, namun tidak melakukan apapun juga merupakan aktivitas yang dapat dilakukan wisatawan (Jang et al., 2002). Apapun yang kegiatan yang dilakukan wisatawan, tetap mereka harus mengeluarkan uang untuk membiayai aktivitas selama liburan. Hal ini yang mendasari mengapa mereka yang bekerja di bidang pariwisata berusaha sedapat mungkin untuk menguasai bahasa sebanyak mungkin, terutama jika kunjungan turis dari negara tersebut jumlahnya signifikan.

Studi pada penelitian terdahulu yang menggunakan metode kualitatif untuk pengelompokan wisatawan tidak dapat dengan jelas menunjukkan dari negara mana turis tersebut berasal dan hanya menyebutkan kawasan seperti Eropa, Amerika Utara (Uzama, 2012). Meskipun studi telah menunjukkan bahwa turis asing tidak dapat digabungkan menjadi satu kelompok, hasilnya tidak menunjukkan bagaimana kelompok dibentuk berdasarkan berbagai macam pengeluaran yang dilakukan oleh wisatawan. Sehingga berdasarkan pada *research gap* yang telah diuraikan diatas, *novelty* pada penelitian ini melakukan pengelompokan untuk pengeluaran turis. Penelitian ini menggunakan data turis dari berbagai negara yang berkunjung ke Jepang.

Tujuan dari penulisan hasil penelitian ini adalah:

1. Membuat dan mengevaluasi model pengelompokan berdasarkan kategori pengeluaran turis selama di Jepang.
2. Membuat dan mengevaluasi model peramalan jumlah turis manca negara ke Jepang.

2. Tinjauan Pustaka

2.1 Penambangan Data

Penambangan data muncul sejak tahun 1980an dari bidang pembelajaran mesin, statistik, dan basis data, yang kini berkembang menjadi alat penting untuk memperoleh pembelajaran berharga dari informasi yang dikumpulkan dalam basis data organisasi (Bach, Schatten, & Marušić, 2013). Dengan kemajuan teknologi, dan di era revolusi industri 4.0, proses penambangan data menjadi lebih cepat dengan biaya yang lebih terjangkau. Penerapan penambangan data sangat luas dan secara umum hasil pengetahuannya digunakan untuk melakukan tindakan yang dikenal dengan keputusan yang didorong oleh data (*data-driven*). Data yang biasanya dianalisis dalam penambangan data berukuran sangat besar mulai dari

ratusan hingga jutaan seperti data penjualan, media sosial, data transaksi yang dianalisis dengan piranti lunak dan teknik statistik (Bach et al., 2013).

Secara singkat, langkah-langkah dalam penambangan data dapat dijelaskan mulai dari membangun pemahaman aplikasi seperti pemahaman bisnis, masalah dan tujuan (Kurgan & Musilek, 2006). Pada langkah pertama ini, perlu dipelajari tujuan yang ingin dicapai dan pengetahuan yang telah ada sebelumnya (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Setelah itu diperlukan penyaringan data agar data memiliki kualitas yang baik sebelum dilakukan pengolahan. Kualitas data yang baik merupakan hal yang wajib diikuti, dan jika tidak, maka dapat menimbulkan berbagai efek negatif yang mungkin terjadi seperti kegagalan dalam suatu proyek (V. Gudivada & Ding, 2017). Persiapan data merupakan bagian awal yang penting dari keseluruhan langkah untuk memperoleh pengetahuan dalam metodologi penambangan data. Tahap ini dimulai dari tahap pembersihan, integrasi hingga melakukan konversi data menjadi satu format dalam program, baik data numerik maupun teks (Han, Kamber, & Pei, 2011).

Langkah selanjutnya yaitu pemilihan dan menjalankan algoritma penambangan data, menginterpretasikan pola hingga mengidentifikasi pola-pola menarik yang ditemukan berdasarkan pengujian secara empiris. Hasil yang didapat diharapkan merupakan suatu temuan yang bisa dipahami oleh banyak orang (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

2.2. Heatmap

Heatmap merupakan visualiasi yang menggambarkan korelasi antara dua variabel. Korelasi merupakan bagian penting dalam machine learning karena memungkinkan peneliti untuk menemukan hubungan yang dalam suatu data yang berukuran besar (Nilsson 2005). Matriks korelasi adalah matriks kovarians dari variabel acak yang distandarisasi. Nilai koefisien korelasi didapatkan dengan membagi setiap variabel acak dengan standar deviasi dalam matriks korelasi (Deisenroth, Faisal et al. 2020). Nilai korelasi berkisar antara -1 hingga 1, semakin besar angka menuju 1 menunjukkan hubungan linear yang positif diantara dua variabel. Namun jika angka yang didapatkan adalah 0 berarti tidak ada asosiasi diantara 2 variabel tersebut (Montgomery and Runger 2018).

2.3. K-Means Clustering

Untuk mencapai tujuan tersebut, dapat menggunakan k-means, salah satu algoritma pembelajaran tanpa pengawasan (*unsupervised machine learning*) yang paling terkenal untuk memecahkan masalah pengelompokan (Ding & He, 2004). Metodenya yaitu dengan melakukan klasifikasi sekumpulan data melalui sejumlah kelompok dengan menentukan suatu titik pusat secara acak untuk setiap kelompok. Jarak antara tiap titik dengan titik pusatnya dihitung dengan menggunakan rumus jarak *Euclidean*. Tahap berikutnya yaitu memperbaiki kualitas kelompok melalui perhitungan variasi di dalam kelompok dengan cara menghitung jumlah kuadrat kesalahan (*within cluster sum of square- WCSS*) (Han et al., 2011). Algoritma k-means kemudian selanjutnya memindahkan titik pusat dan melakukan perhitungan ulang jarak titik pusat dengan titik-titik di sekitarnya berdasarkan hasil WCSS. Iterasi kemudian dihentikan jika ditemui tidak adanya perubahan nilai. Hal ini berarti lokasi titik pusat sudah tetap dan tidak berubah lagi. Untuk menentukan kualitas kelompok berdasarkan nilai siluet (Sarstedt & Mooi, 2011).

2.4. ARIMA

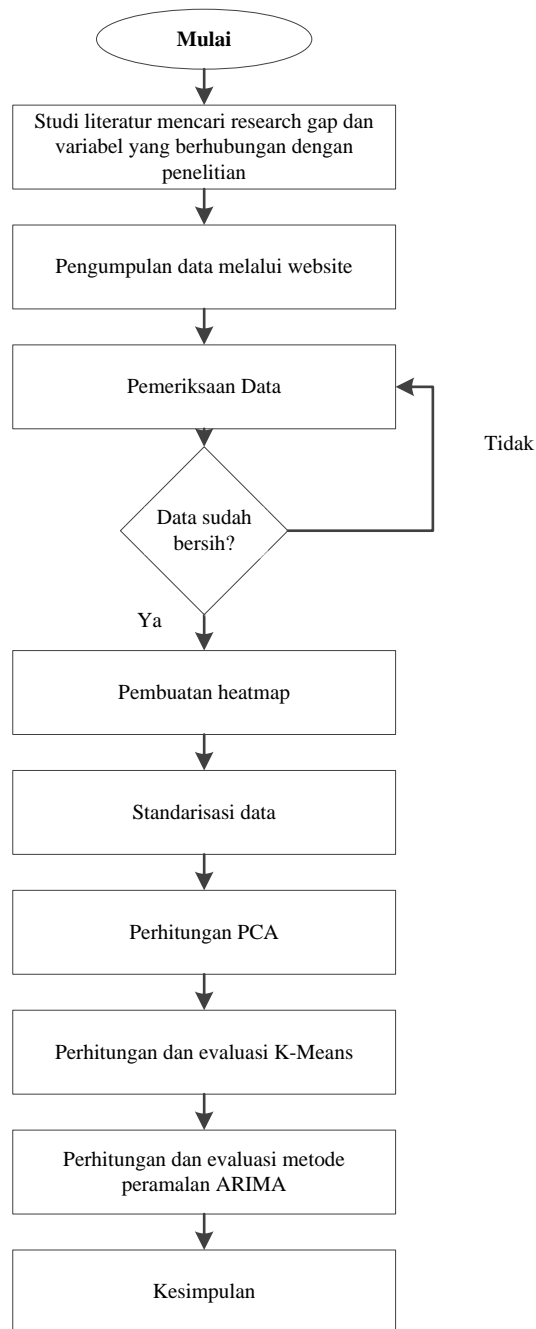
ARIMA merupakan salah satu metode peramalan yang banyak digunakan untuk berbagai macam penelitian seperti aliran Sungai Nil yang menggunakan data bulanan (Montanari, Rosso, & Taqqu, 2000) atau data dalam satuan jam untuk penelitian di bidang teknologi informasi (Tran, Debusschere, & Bacha, 2012). Dalam industri jasa, ARIMA juga dipakai untuk penelitian di bidang pariwisata yang memiliki data musiman (Chen, Chang, & Chang,

2009). Data musiman berfluktuatif tergantung dari berbagai sebab seperti liburan sekolah ataupun musim di suatu negara.

Untuk menangani data musiman, maka data yang dimiliki perlu dibedah untuk melihat komponen pembentuk pola data tersebut. Data musiman memiliki kecenderungan naik turun di periode tertentu dan berulang di tahun berikutnya. Komponen data yang berbentuk seasonal memiliki bagian yaitu tren, musim dan residu (Hillmer & Tiao, 1982). Hasil yang didapatkan mungkin tidak baik, jika ada data yang dipengaruhi musim di dalam model yang digunakan (N.Kulendran & L.King, 1997). Karena itu, musim merupakan elemen yang harus diperhitungkan, namun dihilangkan terlebih dulu untuk melakukan suatu peramalan. Peramalan menggunakan ARIMA membutuhkan *auto correlation function* (ACF), *partial autocorrelation function* (PACF). Nilai ACF menunjukkan autokorelasi yang mengukur hubungan linier antara *lagged value* dari deret waktu, sedangkan PACF mengukur nilai autokorelasi dengan menghilangkan efek dari *lagged value* yang berada diantara dua angka yang diukur (Hyndman and Athanasopoulos 2018). Pemilihan parameter yang terbaik berdasarkan nilai *Akaike Information Criterion* (AIC), digunakan dalam berbagai penelitian mengenai turisme (Elamin and Fukushige 2017, Janjua, Muhammad et al. 2021).

3. Metodologi Penelitian

Gambar 1 menunjukkan langkah-langkah pada penelitian ini:



Gambar 1. Metode Alir Penelitian

- a. Studi literatur ini dilakukan untuk mencari *research gap* dari penelitian para akademisi terkait pariwisata.
- b. Pengumpulan data melalui website
- c. Melakukan pemeriksaan data untuk mencari apakah ada variabel yang hilang atau tidak lengkap. Jika data sudah bersih, maka dapat dilanjutkan ke tahap berikutnya.
- d. Pembuatan *heatmap* untuk melihat korelasi antara satu variabel dengan variabel lainnya.
- e. Melakukan standarisasi data untuk membuat semua variabel yang besarnya bervariasi menjadi satu skala yang sama.
- f. Melakukan perhitungan PCA untuk mereduksi variabel menjadi 2 dan dipetakan ke dalam sumbu x dan y.

- g. Melakukan perhitungan dan evaluasi K-Means dari hasil perhitungan dengan program komputer.
- h. Melakukan perhitungan dan evaluasi peramalan dengan metode ARIMA.
- i. Penarikan kesimpulan dari analisis yang telah dilakukan.

4. Hasil dan Pembahasan

Penelitian ini didasarkan pada penelitian yang telah dilakukan pada penelitian sebelumnya dimana studi ini melakukan pengelompokan negara-negara seperti Eropa dan Amerika Utara, namun tidak menyebutkan dengan jelas asal negaranya (Uzama, 2012). Penelitian ini juga tidak memiliki kategori pengeluaran wisata seperti makanan, transportasi, belanja, dan lain-lain yang dikeluarkan oleh masing-masing wisatawan. Karena itu, jika berbicara tentang strategi marketing, maka sulit untuk melakukan strategi promosi yang harus dilakukan, karena harus jelas negara mana yang menjadi sasaran pemasaran. Untuk melakukan pengelompokan, pengumpulan data dilakukan dengan mengambil data historis dari situs Japan National Tourism Organization (JNTO). Situs ini memiliki data historis yang memadai dengan mencakup hal yang cukup detail seperti jenis pengeluaran dan negara asal turis.

Data historis yang dikumpulkan merupakan data rata-rata pengeluaran per wisatawan mulai tahun 2010 hingga tahun 2019. Data total negara yang mengunjungi Jepang diambil adalah 20 negara yaitu Australia, Kanada, China, Prancis, Jerman, Hongkong, Indonesia, India, Italia, Malaysia, Filipina, Rusia, Singapura, Korea Selatan, Spanyol, Taiwan, Thailand, Inggris Raya, Amerika Serikat, dan Vietnam. Data dari negara lain tidak diambil karena merupakan data agregat Asia dan Eropa yang tidak diklasifikasikan (*unclassified Asia & Europe*). Selain itu juga terdapat data yang tidak lengkap yang tidak dapat dipakai seperti data dari Mongolia, Turki dan Macau.

Para pelancong dari 20 negara tersebut berkontribusi terhadap pendapatan negara Jepang melalui aktivitas selama liburan. Kegiatan mengkonsumsi tersebut dilakukan untuk mendapatkan suatu kesenangan yang dicari oleh para pelancong. Kesenangan termasuk dalam *hedonic value* yang diindikasikan seperti kunjungan yang dilakukan melebihi ekspektasi, program turis menarik dan ada kejutan yang menyenangkan (Calver & Page, 2013). Jadi konsumsi hedonis tidak mencari hal nyata sudah diketahui, tetapi kepada apa yang diharapkan untuk terjadi (Hirschman & Holbrook, 1982). Pada penelitian ini, data terdiri dari 6 variabel berupa rata-rata pengeluaran wisatawan per kapita untuk akomodasi, makanan-minuman, hiburan, lain-lain, belanja dan transportasi.

4.1 Pemeriksaan Data

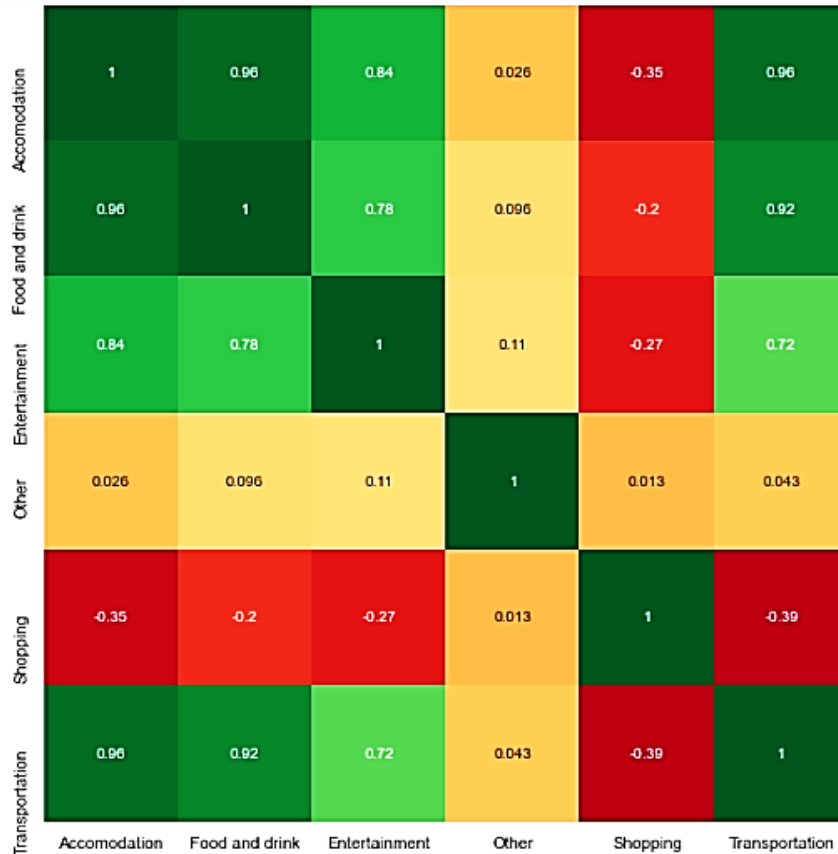
Pada proses awal dilakukan pemeriksaan untuk melihat apakah ada data yang berulang, hilang atau tidak lengkap. Hal ini penting dilakukan untuk memastikan kualitas data yang akan dilakukan untuk tahap selanjutnya. Setelah dilakukan pemeriksaan, ditemukan bahwa data yang dikumpulkan bervariasi jumlahnya dan tidak lengkap untuk tiap negara. Misalnya, data Indonesia untuk kunjungan ke Jepang dimulai pada 2015, sedangkan data Australia dimulai pada 2010 dan data Italia dimulai pada 2015. Karena ukuran data yang berbeda, maka angka yang tidak lengkap tidak dapat diproses dalam program. Untuk pengelompokan dibutuhkan matriks yang lengkap, karena itu data digunakan mulai tahun 2015. Setelah direduksi jumlah barisnya, didapatkan matriks yang berukuran 6 x 20 dan dapat digunakan untuk tahap selanjutnya. Untuk seluruh tahapan pengolahan data yaitu dari pembersihan data hingga perhitungan kualitas model menggunakan bahasa program Python dengan piranti lunak Jupyter.

4.2 Matriks *Heatmap*

Matriks *heatmap* berukuran 6 x 6 pada gambar 2 memetakan korelasi kekuatan korelasi untuk kategori pengeluaran, di mana sumbu x dan y adalah kategori pengeluaran. Korelasi antara

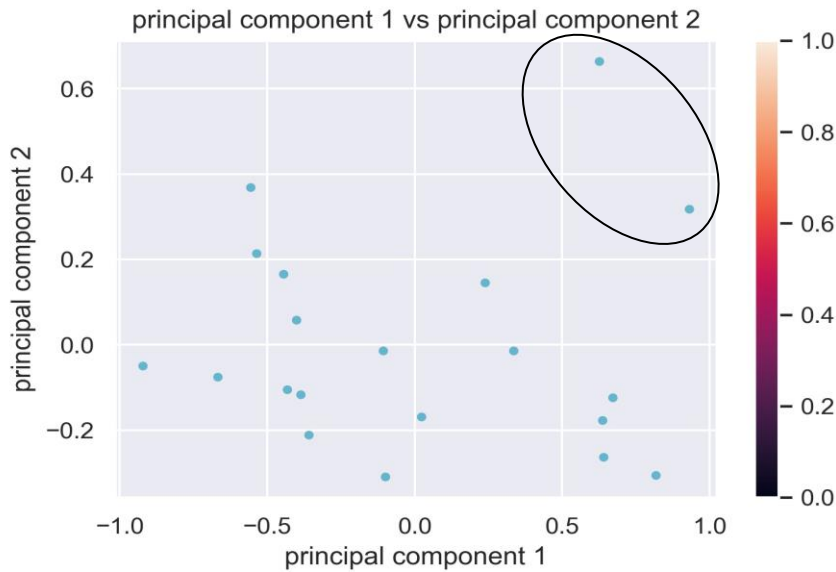
ANALISIS PERAMALAN (Laurence, dkk.)

satu kategori dan kategori lainnya ditunjukkan dengan adanya gradasi warna dari merah ke hijau. Semakin besar korelasi tersebut, maka semakin hijau warnanya. Untuk menjelaskan hal tersebut yaitu turis yang memiliki kemampuan membayar akomodasi mahal juga memiliki kemampuan yang sama untuk mengeluarkan biaya makanan dalam jumlah besar. Hal yang sama juga berlaku untuk kelompok pengeluaran lain seperti makanan & minuman dengan *entertainment*.



Gambar 2. Korelasi kunjungan turis ke Jepang

Pengeluaran turis untuk akomodasi, transportasi dan makanan & minuman, dan hiburan memiliki korelasi yang tinggi. Korelasi negatif ditunjukkan dengan angka variabel belanja dengan transportasi (-0,39), akomodasi (-0,35), makanan & minuman (-0,2), dan hiburan (-0,27). Hal ini menunjukkan bahwa jika turis lebih mengutamakan *shopping*, maka alokasi pengeluaran akan lebih sedikit untuk keperluan transportasi, akomodasi, makanan & minuman dan hiburan. Namun, baik nilai korelasi positif maupun negatif tidak menunjukkan kausalitas, tetapi atribut ini dapat terhubung dengan atribut lain (Han et al., 2011). Dari enam variabel tersebut, semuanya akan digunakan untuk pengolahan data dan tidak ada variabel yang dihilangkan.



Gambar 3. Standarisasi dengan min-max untuk kunjungan turis ke Jepang.

Karena besarnya pengeluaran tiap variabel bervariasi jumlahnya, diperlukan suatu standar angka untuk pengolahan lebih lanjut. Sebagai ilustrasi, pengeluaran untuk akomodasi, angkanya dapat mencapai suatu nilai besarnya berkali lipat dibandingkan pengeluaran untuk makanan-minuman perhari. Proses reduksi dimensi dimulai dengan standarisasi data, menggunakan skala minimum-maksimum, di mana semua angka ditransformasikan sehingga nilai maksimum adalah 1 dan nilai minimumnya adalah -1. Untuk langkah selanjutnya, dari berbagai kategori pengeluaran turis akan direduksi menjadi 2 dimensi yang akan dipetakan ke sumbu x dan y dengan metode *Principal Component Analysis (PCA)*. Pengurangan dimensi merupakan cara untuk mengurangi kompleksitas model dan menghindari *overfitting*. Tujuan pengurangan dimensi ini agar lebih efektif dibandingkan menggunakan ukuran dimensi aslinya (Han et al., 2011). Selain itu jika terlalu banyak data dapat menurunkan akurasi dari model karena data yang perlu digeneralisasi lebih banyak. Dengan cara ini akan didapatkan sumbu baru yang diurutkan sedemikian rupa sehingga sumbu x menunjukkan varians paling banyak di antara data, sumbu y menunjukkan varians tertinggi berikutnya.

Metode PCA merupakan langkah awal yang perlu dilakukan sebelum proses pembuatan kelompok (Ding & He, 2004). Hasil PCA dapat dilihat pada gambar 3, dimana angka telah dipetakan ke sumbu x menjadi komponen utama 1 dan sumbu y sebagai komponen utama 2. Angka-angka tersebut merupakan hasil perhitungan nilai eigen dan vector eigen dengan menggunakan program. Gambar 3 menunjukkan mayoritas titik-titik tersebar berada di daerah kiri dan sebagian kecil menyebar di bagian sebelah kanan. Dari gambar terlihat jarak antara satu titik dengan titik lainnya tidak terlalu dekat antara titik-titik tersebut. Gambar 3 juga menunjukkan perbedaan yang signifikan untuk pengeluaran turis dari negara-negara tersebut yaitu terdapat 2 titik yang letaknya berdekatan di kanan atas, sehingga hal ini menunjukkan pengeluaran turis di Jepang yang berasal dari 2 negara sangat kontras perbedaannya dengan negara lain. Sehingga dugaan awal adalah terdapat pembagian 2 kelompok negara dengan komposisi 18 negara dan 2 negara.

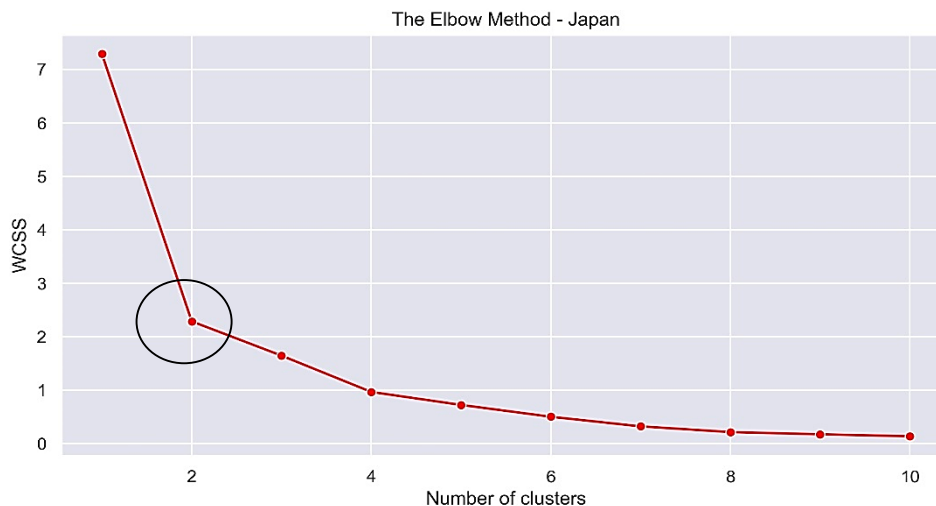
Tabel 1. Proporsi variansi

Komponen Utama	Negara Tujuan
	Jepang
1	68,36%
2	15,48%
Total	83,84%

Titik yang mewakili setiap negara memiliki jarak dengan titik lainnya yang mempengaruhi nilai standar deviasi. Nilai *variance explained* yang didapatkan setelah proses standarisasi adalah 68,36% untuk komponen utama pertama dan 15,48% untuk komponen utama komponen 2. Total kedua komponen tersebut adalah 83,85%. Nilai *variance explained* ini baik dan dapat digunakan untuk proses selanjutnya untuk metode pengelompokan.

4.3 Metode K-Means

Usaha pengelompokan yang dibahas pada bagian sebelumnya yaitu melalui pendugaan awal secara visual mengandung kelemahan yaitu bersifat subjektif dan bervariasi untuk setiap orang yang melihatnya. Metode ilmiah untuk mengatasi persoalan pengelompokan tersebut yaitu metode *k-means clustering*, yang merupakan metode yang populer digunakan dalam pembelajaran mesin. Penentuan jumlah kelompok dengan mencari titik patahan pertama pada kurva. Bentuk dari patahan ini menyerupai bentuk lutut (*knee*) atau siku (*elbow*) pada manusia.



Gambar 4. Lokasi patahan

Gambar 4 menunjukkan adanya patahan pertama yang ditandai dengan tanda lingkaran. Titik tersebut pada sumbu x menunjukkan jumlah kelompok dan sumbu y adalah nilai inersia yaitu *Within-Cluster-Sum-of-Squares* (WCSS). Nilai WCSS sebesar 2,36 dengan jumlah kelompok adalah 2 yang berarti terdapat 2 pengelompokan negara-negara. Walaupun hasil ini sesuai dengan dugaan awal, namun perlu ditinjau anggota dari kelompok tersebut, apakah sama dengan pendugaan awal.

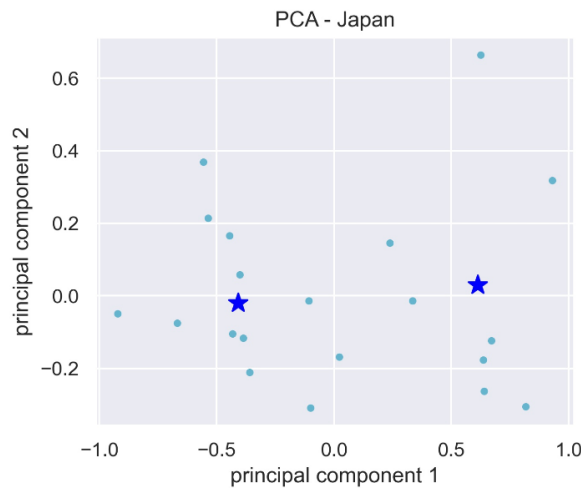
Lokasi titik pusat untuk semua kelompok diperoleh dari algoritma *k-means*. Titik pusat diperoleh setelah melalui tahapan iterasi berulang, dimulai dengan penempatan acak titik pusat sementara awal dan pengukuran jarak antara setiap titik dengan titik tengah. Kemudian titik pusat tersebut dipindahkan dan diukur jaraknya kembali sehingga nantinya akan didapatkan jarak minimum antara titik dan titik pusat dalam kelompok dan jarak maksimum antara titik pusat satu kelompok dengan lainnya (Sarstedt & Mooi, 2011). Dengan demikian pada gambar

akan terlihat jarak yang jauh antara satu titik pusat yang satu dengan titik lainnya sehingga ada pemisahan yang jelas antar kelompok.

Tabel 2. Iterasi untuk menentukan nilai inersia

Iterasi	Nilai Inersia	Titik Pusat 1	Titik Pusat 2
0	2,59	(-0,44; 0,00)	(0,54; 0,00)
1	2,28	(0,61; 0,03)	(-0.40; -0.02)

Untuk menentukan lokasi titik pusat dirancang untuk dilakukan iterasi sebanyak maksimum 100 kali. Jumlah iterasi ini cukup untuk menentukan lokasi titik pusat yang stabil dan tidak berubah. Pada iterasi awal didapatkan nilai inersia sebesar 2,59 dan kemudian turun menjadi 2,28. Hasil ini stabil pada iterasi pertama sehingga tidak perlu mencapai 100 iterasi. Selisih nilai inersia diperoleh dari selisih inersia awal dikurangi inersia akhir yaitu 0,31. Koordinat titik pusat untuk kedua kelompok adalah (0,61; 0,03) dan (-0.40; -0.02) yang ditunjukkan dengan tanda bintang pada gambar 5.



Gambar 5. Lokasi titik pusat

4.4 Evaluasi Kualitas Kelompok

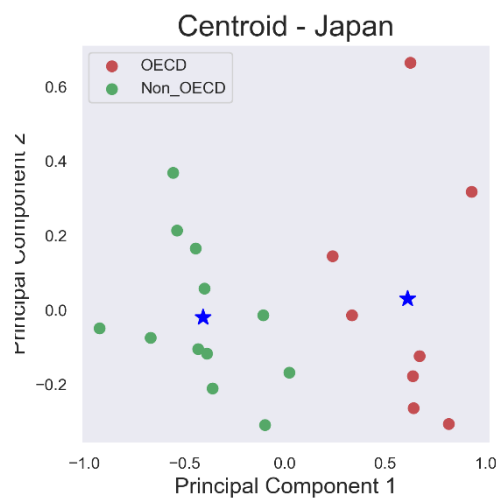
Kualitas kelompok perlu diuji dengan menggunakan skor siluet. Nilai siluet yang berada pada rentang 0,2 hingga 0,5 dikategorikan biasa (Sarstedt & Mooi, 2011), sedangkan nilai yang lebih besar dari 0,5 termasuk dalam kategori baik. Kualitas kelompok dikategorikan buruk jika nilainya di bawah 0,2. Pada penelitian ini dengan menggunakan *scikit* untuk bahasa pemrograman Python didapatkan perhitungan dengan variasi jumlah kelompok untuk mendapatkan nilai siluet yang terbesar. Nilai terbaik yang didapatkan masing-masing adalah 0,56 dimana nilai tersebut lebih besar dari 0,5, sehingga kualitas kelompok dikatakan baik (*Good*). Tabel 3 menunjukkan jumlah kelompok yang lebih dari 2 mengakibatkan skor siluetnya menurun sehingga kualitas kelompok juga berada pada level yang lebih rendah.

Tabel 3. Kualitas kelompok

No	Jumlah Kelompok	Nilai Siluet	Kualitas
1	2	0,56	Baik
2	3	0,41	Biasa
3	4	0,48	Biasa
4	5	0,41	Biasa

ANALISIS PERAMALAN (Laurence, dkk.)

Dengan algoritma pemrograman, sebaran anggota kelompok dapat dilihat pada gambar 5. Secara visual dapat terlihat pembagian yang pengelompokan yang jelas antara 2 gugus tersebut. Tidak terdapat kerancuan dalam hal ini yaitu titik berwarna hijau berada di sekitar titik berwarna merah atau sebaliknya. Jumlah anggota kelompok dan negara dapat dilihat pada tabel 4 yaitu kelompok satu sebanyak 8 negara dan kelompok 2 beranggotakan 12 negara. Untuk penamaan kelompok, jika berdasarkan pendapatan perkapita yaitu tinggi, menengah, dan rendah tidak sesuai dengan hasil perhitungan jumlah kelompok. Kategori lain yang sesuai untuk dipasangkan yaitu pengelompokan apakah negara tersebut merupakan anggota atau non anggota dari *Organization for Economic Co-Operation and Development* (OECD). OECD adalah organisasi yang didirikan dengan tujuan untuk memajukan ekonomi dan perdagangan dunia. *OECD Tourism Committee* yang didirikan pada tahun 1948 secara aktif melakukan berbagai pendekatan untuk mendukung pertumbuhan ekonomi yang berkelanjutan. Turis yang berasal dari negara OECD berkontribusi sebesar kurang lebih 50,5% dari total pengeluaran turis seluruh dunia dalam kurun waktu 2017 hingga 2018 (OECD, 2020). Hasil pengelompokan untuk negara-negara yang tergabung dalam OECD dan non-OECD dapat dilihat pada Gambar 6.



Gambar 6. Negara OECD and non-OECD

Negara yang tergabung dalam OECD adalah yang berada di kelompok 1 yang diwakili oleh titik-titik pada negara-negara hijau dan non-OECD adalah negara-negara yang tergabung dalam kelompok 2 yang diwakili oleh titik-titik merah. Hasil dari pengelompokan ini dapat dijadikan bahan pertimbangan untuk membuat strategi pemasaran yang khusus untuk turis yang berasal negara-negara dalam kelompok 1 dan kelompok 2. Hal ini karena rata-rata GDP per kapita untuk negara 8 negara tersebut dalam kurun waktu 2010 hingga 2019 mencapai 46.337 dolar Amerika, sedangkan rata-rata GDP di kelompok 2 sebesar 26.505 dolar Amerika. Pendapatan per kapita yang besar memungkinkan turis untuk memiliki keleluasaan dan variasi pilihan untuk berbelanja karena ditunjang dengan kemampuan finansial yang tinggi.

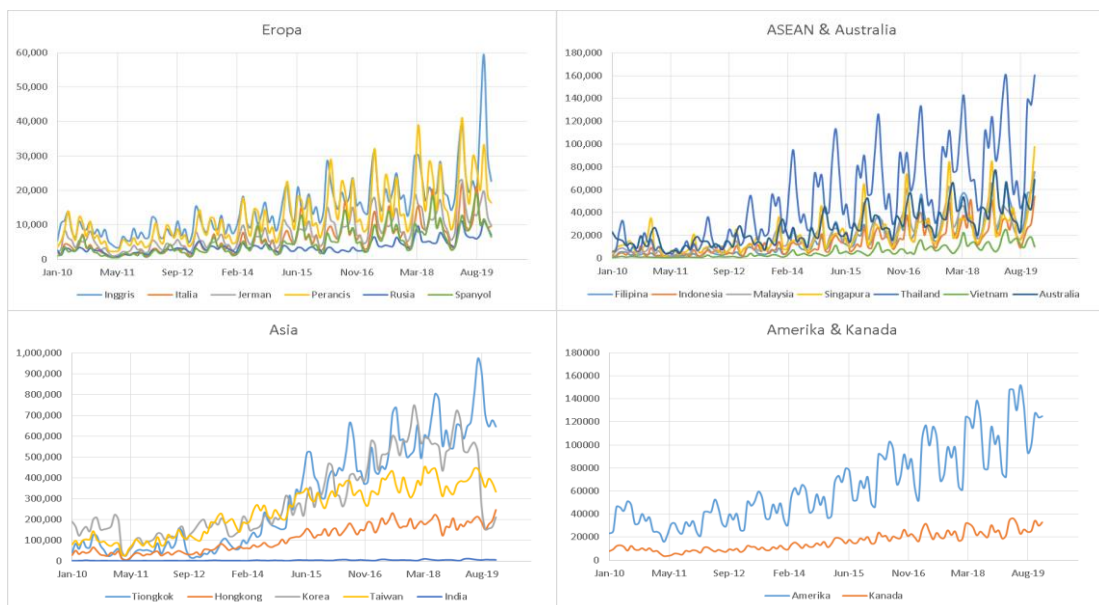
Tabel 4. Negara Asal Turis dan Kelompok

Kelompok	Negara Asal Turis	Jumlah Negara
1	Australia, Kanada, Perancis, Jerman, Italia, Spanyol, Inggris, Amerika Serikat	8
2	Tiongkok, Hongkong, India, Indonesia, Malaysia, Filipina, Rusia, Singapura, Korea Selatan, Taiwan, Thailand, Vietnam	12

4.5 Peramalan

Berdasarkan data kunjungan untuk setiap negara, terdapat kemiripan pola ini dijumpai untuk turis dari berbagai negara. Sejak tahun 2011, kunjungan wisatawan mancanegara mengalami periode yaitu lonjakan yang luar biasa dan hal ini sama untuk seluruh negara tersebut. Gambar 7 menunjukkan garis naik turun yang seperti gigi gergaji menunjukkan adanya pola musiman menunjukkan kenaikan dan penurunan pada musim tertentu. Periode-periode ini seperti musim semi dimana bunga sakura yang sedang mekar menjadi daya tarik tersendiri. Dengan pola data seperti ini maka metode peramalan menggunakan ARIMA perlu ditambahkan faktor musiman, sehingga metode ini tidak murni ARIMA, tetapi berubah menjadi ARIMA musiman (SARIMA).

Tren kunjungan ke Jepang yang meningkat setiap tahun seperti ditunjukkan pada gambar 7 menunjukkan adanya ketertarikan wisatawan untuk mendapatkan pengalaman yang menyenangkan. Secara umum hampir seluruh negara yang terbagi dalam 4 gambar menunjukkan pola yang sama yaitu memiliki tren naik dan pola musiman.

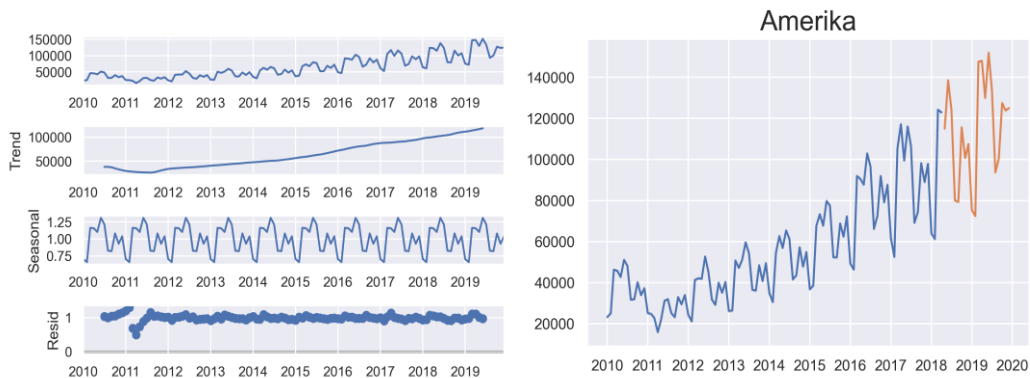


Gambar 7. Kunjungan wisatawan 20 negara ke Jepang

Untuk melihat elemen data diambil data Amerika sebagai contoh seperti dalam gambar 8. Data dapat diuraikan menjadi beberapa komponen yaitu tren, musiman dan sisa nilai. Dengan pola data yang tidak stasioner seperti ini, peramalan tidak dapat dilakukan karena masih terdapat pengaruh tren dan pola musiman di dalam data. Untuk menguji apakah masih terdapat pola musiman maka data diuji dengan tes Augmented Dickey-Fuller (ADF). Jika nilai probabilitas yang dihitung lebih besar dari 0,05, maka data tidak stasioner dan kedua elemen tersebut dihilangkan terlebih dulu dan nanti akan dimasukkan kembali untuk melakukan peramalan.

Untuk membuat model peramalan, data dibagi menjadi dua bagian yaitu data yang dikelompokkan untuk dilatih (*train*) dan diuji (*test*). Jumlah data yang digunakan untuk dimasukkan ke kelompok yang dilatih yaitu berjumlah 100 data pertama. Jumlah ini mewakili 83,3% dari total sampel yang dimiliki. Sisa data yaitu 20 data terakhir akan digunakan untuk dilakukan pengujian. Sebagai ilustrasi, gambar 8 menunjukkan data turis Amerika terbagi menjadi data yang dilatih ditunjukkan dengan garis berwarna biru dan data yang di tes ditunjukkan oleh garis berwarna merah. Untuk perhitungan negara lainnya dilakukan dengan cara yang sama.

ANALISIS PERAMALAN (Laurence, dkk.)



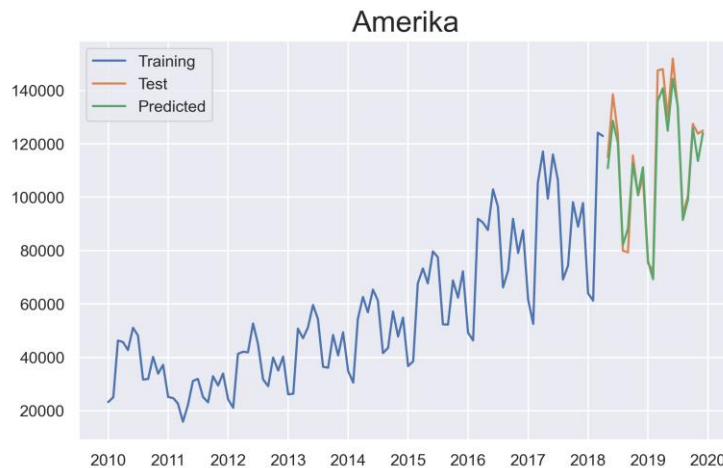
Gambar 8. Data tren, musiman dan residu (kiri) dan data yang dilatih dan dites (kanan).

Untuk melakukan peramalan dengan menggunakan metode ARIMA, umumnya peneliti melihat terlebih dahulu gambar ACF dan PACF dan menentukan pola seperti *dying down*, *cut-off* dan indikator lainnya. Namun dengan fungsi *auto arima* dalam piranti lunak Jupyter, maka ACF dan PACF tidak diperlukan lagi karena algoritma akan mengeluarkan hasil secara otomatis. Tabel 5 menunjukkan hasil iterasi model ARIMA terbaik untuk 20 negara dengan menggunakan *auto arima*. Kombinasi yang dihasilkan terdiri dari dua komponen yaitu tren dan musiman dengan nilai AIC terkecil yang didapat pada iterasi tersebut. Untuk melakukan peramalan dilakukan kombinasi yang dipilih seperti pada tabel dan dihitung nilai AIC (Akaike Information Criteria). Nilai AIC yang didapat untuk pada penelitian ini berada pada jangkauan 1374,38 – 2124,78.

Tabel 5. Kombinasi ARIMA

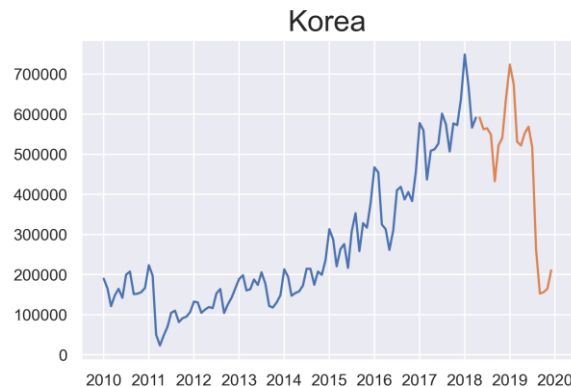
	Negara	Kombinasi	AIC	R ²		Negara	Kombinasi	AIC	R ²
1	Amerika	(0,1,0) (3,1,1)	1746,50	0,94	11	Kanada	(2,1,0)(1,1,0)	1574,90	0,77
2	Australia	(0,1,1) (0,1,1)	1695,14	0,89	12	Korea	(0,1,0)(1,1,1)	2059,56	-2,01
3	Tiongkok	(1,1,0) (2,1,0)	2124,78	0,85	13	Malaysia	(1,1,2)(1,1,2)	1693,53	0,79
4	Filipina	(0,1,1)(3,1,0)	1669,90	0,87	14	Perancis	(2,1,0)(4,1,0)	1567,13	0,95
5	Hongkong	(0,1,0)(3,1,1)	1919,57	0,24	15	Rusia	(0,1,0)(3,1,0)	1374,38	-0,26
6	India	(1,1,1)(1,1,0)	1380,91	0,86	16	Singapura	(0,1,0)(1,1,3)	1680,43	0,98
7	Indonesia	(0,1,1)(0,1,0)	1685,95	0,89	17	Spanyol	(0,1,1)(1,1,0)	1497,42	0,88
8	Inggris	(0,1,2)(2,1,0)	1609,25	0,39	18	Taiwan	(1,1,1)(0,1,2)	1992,66	-0,94
9	Italia	(0,1,1)(1,1,0)	1493,61	0,87	19	Thailand	(0,1,0)(0,1,2)	1806,41	0,88
10	Jerman	(1,1,1)(3,1,1)	1514,35	0,92	20	Vietnma	(0,1,1)(0,1,0)	1467,26	0,83

Data yang sudah dites digunakan untuk melakukan peramalan. Sebagai ilustrasi, hasil pemetaan data yang diuji dan peramalan dapat dilihat pada gambar 9. Garis berwarna hijau dan merah pada gambar 9 berhimpit dan cenderung memiliki pola yang sama dengan garis berwarna biru. Hasil peramalan ini menunjukkan model yang baik dengan indikator R² yang mendekati angka 1.



Gambar 9. Data dilatih, dites, dan diprediksi

Beberapa nilai R^2 pada tabel 4 menunjukkan angka yang lebih kecil dari 0,7 seperti data turis dari Hongkong dan Inggris. Selain itu terdapat nilai R^2 yang negatif yaitu data jumlah turis berkebangsaan Korea Selatan, Rusia dan Taiwan. Nilai R^2 yang kecil maupun negatif ini dapat disebabkan oleh angka yang turun dengan drastis pada akhir periode, seperti pada data Korea Selatan yang ditunjukkan pada gambar 10. Demikian pula halnya dengan negara lainnya, hal tersebut disebabkan pergerakan dari data itu sendiri.



Gambar 10. Data Kunjungan Turis Korea Selatan ke Jepang

Penurunan tersebut disebabkan oleh faktor eksternal yaitu adanya seruan untuk melakukan boikot terhadap produk Jepang pada tahun 2019. Aksi tersebut menimbulkan dampak tidak hanya pada sektor perdagangan saja, tetapi juga berimbas pada sektor pariwisata. Secara umum dapat ditarik kesimpulan bahwa metode ARIMA dapat diterapkan untuk melakukan peramalan kunjungan turis ke Jepang jika tidak terjadi pandemi covid-19. Dampak dari covid-19 yang mulai merebak sejak kuartal pertama tahun 2020 berpengaruh besar terhadap sektor pariwisata dengan adanya pembatasan kegiatan dan perjalanan.

5. Kesimpulan

Data yang distandarisasi dengan metode PCA mereduksi dimensi dari terdiri dari banyak variabel seperti akomodasi, hiburan, makanan-minuman, transportasi, belanja, dan lain-lain menjadi 2 variabel yaitu komponen utama 1 dan komponen utama 2. Data distandarisasi dan ditemukan *variance explained* sebesar 83,84% dan 95,85%. Terdapat dua kelompok yang

terbentuk dari hasil pengolahan data, yang terbagi menjadi negara OECD dan non-OECD untuk data turis tujuan kunjungan ke Jepang. Kualitas kelompok didapatkan dengan menggunakan metode rata-rata siluet dengan nilai masing-masing 0,56 dan 0,67 yang melebihi batas minimal 0,50, sehingga dapat dikategorikan baik.

Dengan data yang dipengaruhi oleh tren dan musim, penggunaan ARIMA yang memasukkan kedua elemen tersebut dapat diterapkan pada 75% data turis mancanegara yang berkunjung ke Jepang. Faktor-faktor eksternal lain yang tidak diketahui ataupun politik yang mempengaruhi hubungan antar negara mampu menjadi faktor pengungkit yang meningkatkan jumlah kunjungan turis atau menjadi penghambat yang mempengaruhi minat turis untuk berkunjung ke suatu negara.

Keterbatasan dalam penelitian ini yaitu belum ditemukannya metode peramalan untuk negara-negara yang memiliki nilai R^2 kecil. Selain itu penelitian ini belum mengikutsertakan data mulai tahun 2020 dimana pandemi covid-19 melanda seluruh dunia. Para praktisi di dunia pariwisata dapat mulai mempersiapkan langkah-langkah yang perlu ditempuh jika pandemi sudah berakhir dan negara-negara membuka kembali pintu masuknya. Jika fase new normal yang saat ini sedang dijalani dan berlanjut ke fase sebelum pandemi, maka pola dan kebiasaan para turis mungkin berulang seperti data historis hingga tahun 2019. Saran penelitian lebih lanjut yaitu meneliti sampai sejauh mana data di masa pandemi dapat digunakan sebagai dasar untuk melakukan peramalan.

6. Daftar Pustaka

Andonian, A., Kuwabara, T., Yamakawa, N., & Ishida, R. (2016) 'The future of Japan's tourism: path for sustainable growth towards 2020' [Online]. Tersedia di: https://www.mckinsey.com/~media/mckinsey/industries/travel_logistics_and_infrastructure/our_insights/can_inbound_tourism_fuel_japans_economic_growth/the_future_of_japans_tourism_full_report.pdf (Diakses: 3 Mei 2021 pada jam 15.48).

Bach, M. P., Schatten, M., & Marušić, Z. (2013) 'Data mining applications in tourism: a keyword analysis', *Central European Conference on Information and Intelligent Systems*, 18-20 September 2003, pp. 26-32.

Calver, S. J., & Page, S. J. (2013) 'Enlightened hedonism: exploring the relationship of service value, visitor knowledge and interest, to visitor enjoyment at heritage attractions', *Tourism Management*, Vol 39, pp. 23-36.

Chen, C.-F., Chang, Y.-H., & Chang, Y.-W. (2009) 'Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan', *Transportmetrica*, Vol 5 Issue 2, pp. 125-140.

Deisenroth, M. P., et al. (2020) *Mathematics for machine learning*. Cambridge University Press.

Ding, C., & He, X. (2004) 'K-means clustering via Principal Component Analysis', *Proceedings of the 21st International Conference on Machine learning*, Juli 2004.

Dwyer, L., & Forsyth, P. (2008) 'Economic measures of tourism yield: what markets to target?', *International Journal of Tourism Research*, Vol. 10(2), pp. 155-168.

Elamin, N. and M. Fukushige (2017) 'Forecasting extreme seasonal tourism demand: the case of Rishiri Island in Japan', *Asia-Pacific Journal of Regional Science*, Vol 2 Issue 2, pp. 279-296.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996) 'Knowledge discovery and data mining: towards a unifying framework', *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Agustus 1996, pp. 82-88.

Frechtling, D. C. (2001), *Forecasting tourism demand. methods and strategies*. Elsevier.

Han, J., Kamber, M., & Pei, J. (2011) *Data mining: concepts and techniques*. Morgan Kaufmann.

Hillmer, S. C., & Tiao, G. C. (1982) 'An ARIMA-model-based approach to seasonal adjustment', *Journal of the American Statistical Association*, Vol 77, pp. 63-70.

Hirschman, E. C., & Holbrook, M. B. (1982) 'Hedonic consumption: emerging concepts, methods and propositions', *Marketing*, Vol 46 Issue 3.

Hyndman, R. J. and G. Athanasopoulos (2018) *Forecasting_ principles and practice*. Otexts.

Janjua, L. R., et al. (2021) 'Impact of COVID-19 pandemic on logistics performance, economic growth and tourism industry of Thailand: an empirical forecasting using ARIMA', *Brazilian Journal of Operations & Production Management*, Vol 18, Issue 2.

Lim, C., & McAleer, M. (2001) 'Monthly seasonal variations: Asian tourism to Australia', *Annals of Tourism Research*, Vol 28, Issue 1, pp. 68-82.

Ministry of Foreign Affairs of Japan, *Exemption of visa (short-term stay)* [Online]. Tersedia pada: https://www.mofa.go.jp/j_info/visit/visa/short/novisa.html (Diakses 3 Mei 2021 pada jam 15.57).

Montanari, A., Rosso, R., & Taquq, M. S. (2000) 'A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan', *Water Resources Research*, Vol 36, Issue 5, pp. 1249-1259.

Montgomery, D. C. and G. C. Runger (2018) *Applied statistics and probability for engineers*. Wiley.

N.Kulendran, & L.King, M. (1997) 'Forecasting international quarterly tourist flows using error-correction and time-series models', *International Journal of Forecasting*, Vol 13(3), pp. 319-327.

Nilsson, N. J. (2005) *Introduction to machine learning*. Stanford University

OECD. (2020), *OECD Tourism trends and policies 2020*.

Sarstedt, M., & Mooi, E. (2011) *A concise guide to market research*. Springer.

Silva, F. J. F., Câmara, G. F. M., Vieira, J. A. C., & Santos, C. A. S. M. (2020) 'Is the spending behaviour of tourists affected by low-cost carriers' operation? Some empirical evidence', *Tourism Management Perspectives*, Vol. 33.

Tafel, M., & Szolnoki, G. (2020) 'Estimating the economic impact of tourism in German wine regions', *International Journal of Tourism Research*, Vol. 22(6), pp. 788-799.

ANALISIS PERAMALAN (Laurence, dkk.)

Tran, V. G., Debusschere, V., & Bacha, S. (2012) 'Hourly server workload forecasting up to 168 hours ahead using Seasonal ARIMA model', *IEEE International Conference on Industrial Technology*, Athens, Greece.

Uzama, A. (2012) 'Yokoso! Japan: classifying foreign tourists to Japan for market segmentation', *Journal of Hospitality Marketing & Management*, Vol. 21(2), pp. 132-154.

V. Gudivada, A. A., & Ding, J. (2017) 'Data quality considerations for big data and machine learning: going beyond data cleaning and transformations', *International Journal on Customer Relations Advances in Software*, Vol. 10(1), pp. 1 - 20.

World Travel & Tourist Council, *Economic impact report* [Online]. Tersedia pada: <https://wttc.org/Research/Economic-Impact>. (Diakses 3 Mei 2021 pada jam 15.48).