

PENERAPAN *HIDDEN MARKOV MODEL* PADA PERAMALAN PELUANG PERPINDAHAN JUMLAH PENGGUNA MERK *SIM CARD* DI KALANGAN MAHASISWA S1 UNIVERSITAS ANDALAS

RISMA YULIA, DODI DEVIANTO, MAIYASTRI

*Jurusan Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus Unand Limau Manis, Padang, Indonesia
email : risma_16yulia@yahoo.co.id*

Abstrak. *Hidden Markov Model* adalah perkembangan dari rantai Markov dimana *state* tidak dapat diamati secara langsung (tersembunyi), tetapi hanya dapat diobservasi melalui suatu himpunan pengamatan lain. Penelitian ini bertujuan untuk memramalkan peluang perpindahan jumlah pengguna merk *SIM card* di kalangan mahasiswa S1 Universitas Andalas angkatan 2014 dan 2015 pada tahun 2017 dalam selang waktu satu minggu dengan menggunakan algoritma *Baum-Welch* dalam *Hidden Markov Model* dan untuk memprediksi *state* tersembunyi atau peluang naik, turun atau tetapnya jumlah pengguna merk *SIM card* Telkomsel atau Indosat pada tahun 2017 dengan menggunakan *Decoding Problem*. Penelitian ini menggunakan data perpindahan jumlah pengguna merk *SIM card* dengan periode satu minggu. Data yang digunakan yaitu data primer dengan menggunakan kuesioner yang dibagikan kepada responden. Dari hasil penelitian menunjukkan bahwa *Hidden Markov Model* dapat digunakan untuk meramalkan peluang perpindahan jumlah pengguna merk *SIM card* di kalangan mahasiswa S1 Universitas Andalas paling tinggi untuk periode satu minggu kedepan yaitu pada *SIM card* merk Telkomsel ke Indosat dengan menggunakan algoritma *Baum-Welch*. Pada algoritma *Viterbi*, dapat diambil kesimpulan bahwa untuk tahun 2017 kemungkinan peluang perpindahan jumlah pengguna merk *SIM card* yang berpindah dari suatu *SIM card* merk Telkomsel ke Indosat dan Indosat ke Telkomsel di kalangan mahasiswa S1 Universitas Andalas adalah sama.

Kata Kunci: Hidden Markov Model, Decoding Problem, state, algoritma Viterbi, algoritma Baum-Welch, SIM card

1. Pendahuluan

Ada beberapa perusahaan telekomunikasi di Indonesia yang memasarkan beberapa *SIM card*. Masing-masing *SIM card* tersebut mempunyai keunggulan dan kekurangan. Faktor-faktor yang ada antara lain: daya jelajah, layanan data, tarif dan lain sebagainya. Dalam penentuan *SIM card* mana yang digunakan, tentunya disesuaikan dengan kebutuhan pelanggan. Namun dari beberapa faktor yang ada, tariflah yang banyak dijadikan patokan pemilihan *SIM card*. Akibatnya perang tarif pun terjadi.

Untuk menarik pelanggan, banyak penawaran yang diberikan oleh perusahaan telekomunikasi, antara lain: sms gratis, telepon gratis, internetan gratis sampai dengan adanya kartu perdana yang harganya lebih murah dari nilai pulsa yang

dikandungnya. Kartu perdana murah tersebut semula digunakan untuk menarik pelanggan, justru menimbulkan masalah baru. Setelah nilai pulsa dalam kartu habis, konsumen akan membuang kartu tersebut dan membeli kartu perdana baru dan begitu seterusnya, sehingga kartu perdana tersebut hanya dijadikan sebagai kartu panggil (*calling card*) saja.

Oleh karena itu sering mengganti kartu yang juga di kenal sebagai fenomena *churn* sudah seharusnya mendapat perhatian dari perusahaan telekomunikasi karena semakin ketat persaingan dalam bisnis seluler semakin besar *churn rate* yang di timbulkannya. Pada fenomena *churn*, karena keputusan seorang konsumen untuk mengganti kartu prabayarnya tidak di pengaruhi oleh konsumen yang lain atau pun oleh *SIM card* yang digunakan sebelumnya, maka kejadian tersebut termasuk kejadian yang independen, sehingga dapat dimodelkan dalam rantai Markov.

Rantai Markov adalah rangkaian proses kejadian dimana peluang bersyarat kejadian yang akan datang tergantung pada kejadian sekarang. Analisis Markov merupakan suatu bentuk khusus dari model probabilistik yang lebih umum dikenal dengan proses stokastik. Konsep dasar analisis Markov sendiri adalah keadaan dari sistem atau keadaan transisi yang disebut dengan *state*. Sifat dari proses ini adalah apabila diketahui proses berada dalam suatu keadaan tertentu, maka peluang berkembangnya proses dimasa yang akan datang hanya tergantung pada keadaan saat ini dan tidak tergantung pada keadaan sebelumnya [2]. Pasangan kejadian dan penyebab kejadian yang tidak dapat diamati (*hidden*) dan membentuk rantai Markov disebut *Hidden Markov Model (HMM)*.

Hidden Markov Model (HMM) adalah perkembangan dari rantai Markov dimana *state* tidak dapat diamati secara langsung (tersembunyi), tetapi hanya dapat diobservasi melalui suatu himpunan pengamatan lain. Setelah perkembangannya diakhir tahun 1970, *Hidden Markov Model* telah terbukti sangat kuat dan fleksibel dalam pemodelan statistik yang menggambarkan berbagai jenis data terurut. *Hidden Markov Model* terdiri atas sebuah sinyal yang dimodelkan sebagai sebuah rantai Markov keadaan terhingga dan sebuah observasi yang dimodelkan sesuai proses observasi pada rantai Markov, hanya saja proses stokastik pada *Hidden Markov Model* merupakan proses stokastik ganda dimana salah satu prosesnya tidak dapat diamati (*hidden*) [1]. Proses (*hidden*) ini hanya dapat diamati melalui proses yang dapat di observasi. Jika $X = \{X_1, X_2, \dots\}$ adalah sebuah proses Markov dan $O = \{O_1, O_2, \dots\}$ adalah fungsi dari X , maka X adalah sebuah *Hidden Markov Model* yang dapat diobservasi melalui O atau dapat ditulis untuk suatu fungsi f .

Parameter X menyatakan *state process* yang tersembunyi (*hidden*), sementara parameter O menyatakan ruang observasi yang dapat diobservasi. Ciri-ciri *Hidden Markov Model* adalah sebagai berikut.

- (a) Observasi diketahui tetapi urutan keadaan (*state*) tidak diketahui sehingga disebut (*hidden*).
- (b) Observasi adalah fungsi probabilitas keadaan.
- (c) Perpindahan keadaan adalah dalam bentuk probabilitas.

Elemen-elemen dari *Hidden Markov Model* adalah:

- (a) Banyaknya elemen-elemen keadaan tersembunyi (*hidden state*) yang dinyatakan dengan N sebagai jumlah *state* dimana nilai peluang per ruang *state* dinyatakan dengan $S = \{S_1, S_2, \dots, S_N\}$.
- (b) Jumlah pengamatan observasi tiap *state* yang dinyatakan dengan M , dimana peluang setiap keadaan dinyatakan dengan $v = \{v_1, v_2, \dots, v_M\}$ dan ruang observasi dinyatakan dengan $O = \{O_1, O_2, \dots, O_T\}$, dimana T merupakan panjang dari data observasi.
- (c) Matriks peluang transisi, $A=[a_{ij}]$, dimana a_{ij} adalah elemen dari A yang merupakan peluang bersyarat dari keadaan pada saat $t + 1$, jika diketahui keadaan X pada saat t atau

$$a_{ij} = P(X_{t+1} = j | X_t = i).$$

- (d) Distribusi peluang observasi pada saat t , pada keadaan j , yang biasa dikenal dengan matriks emisi $B = [b_{ik}]$, dimana

$$b_{ik} = P(O_t = V_k | X_t = i).$$

- (e) Distribusi *state* awal yang dinyatakan dengan $\pi(i)$, dimana

$$\pi(i) = P(X_1 = i).$$

Hidden Markov Model dapat dituliskan dalam notasi $\lambda = (A, B, \pi)$, dimana A dinyatakan dengan matriks peluang transisi, B merupakan matriks peluang observasi atau dikenal sebagai matriks emisi dan π adalah distribusi keadaan *state* awal.

2. Masalah-masalah Utama dalam *Hidden Markov Model (HMM)*

Terdapat tiga masalah utama yang harus di pecahkan agar *Hidden Markov Model (HMM)* dapat digunakan dalam suatu aplikasi nyata. Masalah tersebut adalah:

2.1. Menghitung Peluang Observasi (Evaluation Problem)

Bila diketahui sebuah model $\lambda = (A, B, \pi)$ dan sebuah barisan observasi $O = \{O_1, O_2, \dots, O_T\}$, kemudian akan dihitung $P(O|\lambda)$ yang dapat ditulis sebagai berikut:

$$P(O|\lambda) = \sum_x P(O|X, \lambda)P(X|\lambda),$$

dimana $X = \{X_1, X_2, \dots\}$ adalah suatu barisan $P(O|X, \lambda)$ adalah probabilitas barisan observasi O untuk suatu barisan *state* X dan $P(X|\lambda)$ merupakan probabilitas dari X bila diberikan sebuah model. Karena *Hidden Markov Model* barisan observasi diasumsikan *independent*, maka

$$P(O|X, \lambda) = \prod_{t=1}^T P(O_t | X_t, \lambda) = b_1(O_1) b_2(O_2), \dots, b_T(O_T),$$

$$P(X|\lambda) = \pi(1) a_{12} a_{23} \dots a_{T-1, T},$$

sehingga diperoleh,

$$P(O|\lambda) = \sum_x P(O|X, \lambda)P(X|\lambda),$$

$$= \sum_{1,2,\dots,T} \pi(1) a_{12} b_1(O_1) b_2(O_2) a_{23} \dots b_T(O_T) a_{T-1, T}.$$

untuk menghitung $P(O|\lambda)$ diperlukan algoritma maju dan algoritma mundur.

Pertama, menghitung peluang observasi dengan menggunakan algoritma maju (*The Forward Algorithm*). Algoritma ini adalah proses iterasi yang didasarkan pada perhitungan peluang bersyarat melalui sifat-sifat pada peluang. Dengan menggunakan definisi peluang bersyarat $P(O|\lambda)$ dapat dihitung, namun operasi perhitungan yang dibutuhkan akan bertambah panjang barisan observasi yang ada, diasumsikan variabel maju $\alpha_t(i)$ didefinisikan sebagai:

$$\alpha_t(i) = P(O_1, O_2, \dots, O_T, X_T = i | \lambda),$$

yaitu peluang barisan observasi O_1, O_2, \dots, O_T dan *state* i pada waktu t dimana $t = 1, 2, \dots, T$ jika diberikan model.

Secara umum algoritma maju terdiri dari tiga bagian, yaitu:

- (a) Tahap inisialisasi
Mencari nilai $\alpha_1(i) = \pi(i)b_i(O_1)$
- (b) Tahap induksi
Mencari nilai $\alpha_{t+1}(j) = \{\sum_{i=1}^N \alpha_t(i)a_{ij}b_j(O_{t+1})\}$
- (c) Tahap terminasi
Mencari nilai $P(O|\lambda) = \sum_{i=1}^N \alpha_T(i)$

Kedua, menghitung peluang observasi dengan menggunakan algoritma mundur (*The Backward Algorithm*). Langkah algoritma mundur hampir sama dengan algoritma maju. Namun bedanya, pada algoritma mundur inisialisasi didasarkan pada seluruh observasi yang ada. Jadi dengan cara yang sama dapat didefinisikan variabel mundur $\beta_t(i)$ sebagai berikut:

$$\beta_t(i) = (O_{t+1}, O_{t+2}, \dots, O_T | X_t = i, \lambda),$$

yaitu kemungkinan rangkaian pengamatan dari $t+1$ hingga T jika diberikan *state* i pada saat t dan model yang diberikan. $\beta_t(i)$ dapat diselesaikan sebagaimana $\alpha_t(i)$.

Secara umum algoritma mundur terdiri dari tiga bagian, yaitu:

- (a) Tahap inisialisasi
Menetapkan $\beta_t(i) = 1$
- (b) Tahap induksi
Mencari nilai $\beta_t(i) = \{\sum_{j=1}^N b_j(O_{t+1})a_{ij}\beta_{t+1}(j)\}$
- (c) Tahap terminasi
Mencari nilai $P(O|\lambda) = \sum_{i=1}^N b_i(O_t)\pi(i)\beta_t(i)$

Parameter-parameter tersebut diantaranya akan digunakan untuk menentukan matriks peluang yang digunakan untuk menghitung peramalan perpindahan jumlah pengguna merk *SIM card* pada periode yang akan datang.

2.2. Menentukan Barisan Keadaan Tersembunyi (Decoding Problem)

Permasalahan kedua pada *Hidden Markov Model* adalah *Decoding Problem*, yaitu menemukan barisan *state* terbaik (optimal) yang beresilasi dengan observasi O

dari sebuah model λ yang juga telah diketahui [1]. Barisan *state* yang optimal didefinisikan sebagai barisan *state* yang mempunyai probabilitas tertinggi dalam menghasilkan barisan observasi yang telah diketahui sebelumnya. Sehingga pada akhirnya diperoleh suatu barisan *state* X yang akan memaksimumkan $P(O|X, \lambda)$. Misal didefinisikan $\delta_t(i)$ dimana $\delta_t(i) = P(X_t = i|O, \lambda)$. Jika $\delta_t(i)$ dijumlahkan terhadap i , karena $X_t(i)$ merupakan partisi dari X maka menurut aturan *Bayes* mengenai partisi, hasilnya menjadi:

$$\sum_{i=1}^N \delta_t(i) = P(X_t = i|O, \lambda) = 1.$$

Sehingga bisa dinyatakan bahwa *state* yang paling optimal untuk masing-masing t bisa diperoleh dari $X_t^* = \arg \max_{1 \leq i \leq N} \delta_t(i)$. Dengan demikian akan dihasilkan barisan *state* yang paling optimal yaitu, $X_t^* = X_1^*, X_2^*, \dots, X_T^*$ untuk suatu observasi $O = \{O_1, O_2, \dots, O_T\}$ yang diberikan. Namun pencarian barisan *state* yang paling optimal dengan cara tersebut, berpeluang menghasilkan barisan yang tidak valid, karena tidak mempertimbangkan probabilitas transisi *state*. Karena itu, untuk menghindari masalah tersebut perlu digunakan suatu metode yang mempertimbangkan probabilitas transisi *state* pada proses pencarian barisan *state* yang paling optimal. Metode yang banyak digunakan untuk menyelesaikan masalah ini adalah algoritma *Viterbi* [1].

Algoritma *Viterbi* dikenalkan oleh Andrew pada tahun 1967. Algoritma ini pertama kali digunakan untuk menyelesaikan masalah pengkodean yang rumit, namun seiring perkembangan ilmu pengetahuan algoritma *Viterbi* digunakan dalam *Hidden Markov Model* untuk mencari barisan keadaan tersembunyi yang paling optimal dari suatu barisan yang terobservasi [1].

Adapun langkah-langkah dari algoritma *Viterbi* yaitu:

- (a) Tahap inisialisasi
Mencari nilai $\delta_1(i) = b_i(O_1)\pi(i)$ saat $t=1$. Pada tahap ini, nilai $\psi_1(i) = 0$
- (b) Tahap rekursi
Mencari barisan maksimum dari nilai maksimum saat waktu t . $\delta_t(j) = b_j(O_t) \max_{1 \leq i \leq N} \{a_{ij}\delta_{t-1}(i)\}$
- (c) Tahap terminasi
Menentukan $P_* = \max_{1 \leq i \leq N} \{\delta_T(i)\}$,

$$X_T^* = \arg \max_{1 \leq i \leq N} \{\delta_T(i)\}.$$

- (d) Tahap Backtracking
Mencari barisan optimal dari hasil tahap rekursi

$$X_t^* = \psi_{t+1}(X_{t+1}^*) \text{ untuk } t = T - 1, T - 2, \dots, 1.$$

2.3. Menaksir Parameter-parameter Hidden Markov Model (Learning Problem)

Permasalahan ketiga adalah masalah optimasi dan permasalahan yang harus dipecahkan adalah mengestimasi model terbaik yang dapat menjelaskan suatu barisan observasi. Untuk menyelesaikan masalah terakhir dalam *Hidden Markov Model*

ini, biasanya digunakan algoritma *Baum-Welch*. Dimana permasalahan ketiga ini adalah mengubah parameter *Hidden Markov Model*, $\lambda = (A, B, \pi)$ sehingga $P(O|\lambda)$ menjadi maksimal. Dalam algoritma *Baum-Welch*, juga didefinisikan empat variabel yaitu: variabel *forward* (maju), variabel *backward* (mundur), variabel $\xi_t(i, j)$ dan variabel $\gamma_t(i)$ [1].

3. Metode Penelitian

3.1. Data Penelitian

Data yang digunakan pada penelitian dalam Tugas Akhir ini adalah data perpindahan merk *SIM card* yang digunakan oleh mahasiswa Universitas Andalas, sebagai contoh dalam penelitian ini adalah mahasiswa angkatan 2014 dan 2015. Data diperoleh berdasarkan hasil kuesioner yang dikumpulkan pada tanggal 05 Agustus 2017 sampai dengan tanggal 09 Desember 2017.

3.2. Metode Analisis

Pada Tugas Akhir ini perpindahan merk *SIM card* yang digunakan oleh mahasiswa Universitas Andalas dilakukan dengan menggunakan metode *Hidden Markov Model*. Pada pengolahan data menggunakan HMM, tidak dapat dilakukan pengamatan data secara langsung (*hidden*) artinya pada data yang terobservasi pengamatan dilakukan secara tidak langsung.

4. Hasil dan Pembahasan

4.1. Data Perpindahan Jumlah Pengguna Merk SIM Card

Pada penelitian ini, digunakan data perpindahan SIM card hanya pada merk Telkomsel ke Indosat dan perpindahan merk SIM card dari Indosat ke Telkomsel dalam selang waktu satu minggu.

4.2. Evaluation Problem dengan Algoritma Forward dan Backward

Pada Algoritma *Forward* ada tiga tahapan yang harus diselesaikan yaitu:

- (a) Tahap Inisialisasi
 Pada Tahap inisialisasi akan ditentukan variabel maju $\alpha_1(i)$, dimana ada 3 buah α yaitu masing-masing untuk naik(1), turun(2), tetap(3) yaitu kemungkinan rangkaian pengamatan yang berada pada state i pada saat t dimana $t=1$ jika diketahui suatu barisan observasi yaitu Telkomsel ke Indosat. Sehingga diperoleh $\alpha_1(1) = 0,0995$, $\alpha_1(2) = 0,07685$ dan $\alpha_1(3) = 0,08528$.
- (b) Tahap Induksi
 Pada tahap ini akan dihitung $a_2(j)$ dan $a_3(j)$, dimana j masing-masing untuk naik (1), turun (2), tetap (3), yaitu kemungkinan rangkaian pengamatan yang berada pada state j pada saat t dimana $t=2$ dan 3 jika diketahui suatu barisan observasi yaitu Indosat ke Telkomsel dan Telkomsel ke Indosat.

(c) Tahap Terminasi

Pada tahap ini adalah menjumlahkan semua peluang gabungan dari observasi dan Hidden state bila diketahui sebuah model sehingga menghasilkan penyelesaian sebagai berikut.

$$P(O = \text{Telkomsel} - \text{Indosat}, \text{Indosat} - \text{Telkomsel} | \lambda) = 0,04.$$

Pada Algoritma *Backward* juga terdapat 3 tahapan yang harus diselesaikan yaitu:

(a) Tahap Inisialisasi

Pada Tahap inisialisasi akan ditentukan variabel maju $\beta_3(i)$, dimana ada 3 buah β yaitu masing-masing untuk naik(1), turun(2), tetap(3) yaitu kemungkinan rangkaian pengamatan dari $t+1$ hingga T jika diberikan state i pada saat t dan model λ . $\beta_3(1) = \beta_3(2) = \beta_3(3) = 1$

(b) Tahap Induksi

Pada tahap ini akan ditentukan variabel mundur $\beta_t(i)$.

(c) Tahap Terminasi

Pada tahap ini akan dijumlahkan semua peluang gabungan dari observasi dan *hidden state* bila diketahui sebuah model sehingga menghasilkan penyelesaian sebagai berikut.

$$P(O = \text{Telkomsel} - \text{Indosat}, \text{Indosat} - \text{Telkomsel} | \lambda) = 0,04.$$

Berdasarkan hasil dari algoritma mundur (*backward*) diperoleh konsisten dengan solusi yang diperoleh dari algoritma maju (*Forward*) yaitu peluang observasi sebesar 0,04.

4.3. Decoding Problem dengan Algoritma Viterbi

Untuk permasalahan Decoding Problem ini adalah bagaimana menentukan barisan hidden state yang optimal dalam hal ini yaitu naik, turun atau tetapnya jumlah pengguna suatu *SIM card*, sesuai dengan barisan observasi yang telah diasumsikan. Algoritma *Viterbi* terdiri atas empat tahap, diantaranya sebagai berikut:

(a) Tahap Inisialisasi

Pada tahap ini nilai untuk Telkomsel - Indosat yang bersesuaian dengan nilai maksimum dari keadaan tersembunyinya adalah $\psi_1(1) = \psi_1(2) = \psi_1(3) = 0$.

(b) Tahap Rekursi

Pada tahap rekursi ini akan ditentukan peluang terbesar sepanjang t observasi kedua hingga akhir pada state i .

(c) Tahap Terminasi

Pada tahap terminasi ini akan ditentukan peluang terbesar sepanjang t observasi akhir pada state i yaitu $P^* = 0,00444$ dan $X_3^* = 1(\text{naik})$.

(d) Tahap Backtracking

Pada tahap Backtracking ini akan ditentukan barisan state terbaik. Berdasarkan perhitungan pada tahap sebelumnya maka diperoleh barisan

state terbaik, yaitu sebagai berikut: $X_2^* = 1(\text{naik})$, $X_1^* = 2(\text{turun})$ yang artinya barisan naik, turun atau tetapnya jumlah pengguna SIM card yang berpindah dari satu *SIM card* ke *SIM card* lain yang paling mungkin untuk tahun 2017 adalah sama.

4.4. Learning Problem dengan Algoritma Baum-Welch

Untuk menghitung prediksi parameter HMM dengan menggunakan algoritma *Baum-Welch* dapat didefinisikan sebuah variabel baru $\xi_t(i, j)$, yaitu peluang proses yang berada pada state i pada saat t dan berada pada state j pada waktu $t + 1$.

5. Kesimpulan

Berdasarkan penelitian mengenai peramalan peluang jumlah pengguna merk *SIM card* yang berpindah dari *SIM card* merk Telkomsel ke Indosat dan jumlah pengguna merk *SIM card* yang berpindah dari *SIM card* merk Indosat ke Telkomsel pada periode satu minggu di kalangan mahasiswa S1 Universitas Andalas tahun 2017 dengan menggunakan *Hidden Markov Model* dapat disimpulkan bahwa:

- (1) Pada *learning problem* dengan algoritma Baum-Welch kemungkinan peluang jumlah pengguna merk *SIM card* yang berpindah paling tinggi pada periode satu minggu ke depan di kalangan mahasiswa S1 Universitas Andalas yaitu pada *SIM card* merk Telkomsel ke Indosat yaitu sebesar 0,546. Peluang jumlah pengguna merk *SIM card* yang berpindah paling rendah pada periode satu minggu ke depan di kalangan mahasiswa S1 Universitas Andalas yaitu pada *SIM card* merk Indosat ke Telkomsel sebesar 0,675.
- (2) Untuk *decoding problem* dengan menggunakan algoritma *Viterbi* dapat diambil kesimpulan bahwa untuk tahun 2017 kemungkinan peluang jumlah pengguna *SIM card* merk Telkomsel ke Indosat dan Indosat ke Telkomsel di kalangan mahasiswa S1 Universitas Andalas adalah sama.

Daftar Pustaka

- [1] Dymarski, P. 2011. *Hidden Markov Models Theory and Applications*. Intech, Croatia.
- [2] Frederic, P. Miller, Agnes, F. Vandome and MCBrewster, John. 2010. *Hidden Markov Model: Markov Process, Markov Chain, Time, Speech Recognition, Handwriting Recognition, Gesture Recognition, Part-of-speech Tagging, Sheet Music, Partial Discharge, Bioinformatics, Bayesian Inferensia*
- [3] Walter, Zucchini, Lain, L. MacDonald and Roland Langrock. 2016. *Hidden Markov Models For Time Series*. Second Edition. CRC Press, United States.