Pemahaman Kejadian Unik dengan Probabilitas

Aurelia H B Matondang-13510023
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13510023@std.stei.itb.ac.id

Makalah ini berisi pembahasan implementasi dari teori-teori Probabilitas yang telah diajarkan selama senester ganjil tahun ajaran 2012/2013 dalam mata perkuliahan Struktur Diskrit . Pembahasan yang diambil dalam makalah ini adalah Pembahasan dalam masalah ulang tahun. Bagaimana dalam jumlah n orang (yang dipilih secara acak) ada beberapa pasang orang yang memiliki tanggal lahir yang sama.Hal ini dapat dijelaskan dengan beberapa teorema dan ketentuan yang pada akhirnya dapat menjawab (kurang lebih) dari masalah yang dihadapkan lewat makalah ini.

Kata Kunci- Fungsi Hash,"Birthday Attack","Pigeonhole Principle,rantai Markov, Revolving door.

I. PENDAHULUAN

Istilah "kebetulan" sering kali membuat kita merasa terkejut dan memberikan pemikiran-pemikiran untuk mengkaitkan kejadian tersebut ke dalam hal-hal supernatural dan tidak wajar. Hal ini memang tidak bisa dijelaskan langsung oleh sebuah teorema dalam probabilitas. Namun pada kenyataannya, kejadian ini dapat dijelaskan oleh sebuah cabang pengetahuan dari matematika yang mengkaji kasus kebetulan ini. Kasus tersebut adalah kasus terkait tanggal lahit/ulang tahun.

Dari sekian jumlah penduduk di suatu daerah yang dilahirkan setiap hari dan setiap detiknya, pasti terdapat beberapa kasus dimana beberapa dari manusia tersebut memiliki waktu (tanggal lahir) yang sama. Kelahiran tersebut dapat didata dan dapat di prediksikan berapa persen kemungkinan kejadian tersebut terjadi dalam suatu daerah dengan jumlah penduduk yang dikaji dalam jumlah tertentu.

Perhitungan mengenai bagaimana mengetahui seberapa besar kemungkinan tersebut dapat dicari berdasarkan beberapa teori dan hukum tertentu.

Pada bagian teori dan perhitungan akan terdapat salah satu teori yang dapat menjelaskan dan menyelesaika maslah tanggal lahir ini. Namun, untuk melihat permasalahan tanggal lahir ini dalam kasus lainnya, pada bagian contoh kasus lainnya dijelaskan penyelesaian kasus ini dalam sebuah rantai yang disebut rantai markov. Rantai markov ini akan dijelaskan dengan lansung melihat contoh kasus yang telah terjadi di suatu daerah di Amerika dan bagaimana para matematikawan melihat kejadian ini sebagai salah satu misteri matematika yang dapat dipecahkan.

II. BIRTHDAY PROBLEM

Dalam birthday problem ini yang akan dibahas adalah kemungkinan beberapa pasang orang bisa memiliki tanggal lahir yang sama. Contohnya dalam 24 orang yang dipilih secara acak, dengan membandingkan tanggal ulang tahun orang pertama dari 24 orang yang dipilih tersebut dengan 23 orang lainnya, orang pertama tersebut memiliki 23 kesempatan untuk menemukan kesamaan dalam tanggal lahir. Sedangkan orang ke dua memiliki kesempatan 22 untuk menemukan kesamaan tanggal ulang tahun dengan lainnya. Secara umum, untuk menemukan beberapa kemungkinan terbentuknya pasangan (tanpa memperdulikan tanggal ulang tahun dan beberapa hal lain yang dapat mempengaruhi perhitungan probabilitas) dapat dilakukan dengan mencari kombinasi dari jumlah orang yang dipilih secara acak dengan jumlah orang yang akan dipasangkan (2 orang).

$$_{24}C_2 = (24.23)/(2) = 276$$

Untuk memilih sebuah tanggal untuk dibandingkan dengan tanggal lahir orang lainnya dalam 24 orang tersebut, terdapat 1/365 kemungkinan yang dimiliki oleh tanggal lahir dari setiap orang tersebut (dengan tidak memperhitungkan tanggal 29 Febuari). Perhitungan ini membuat perhitungan kemungkinan adanya tanggal lahir yang sama dari 24 orang tersebut secara statistik tidak ekuivalen.

Beberapa kemungkinan dapat terjadi dalam persoalan tanggal lahir ini. Matematikawan Bloom, Clevenson, dan Watkins menunjukkan bahwa probabilitas dari dua tanggal lahir yang sama dari sebuah grup yang terdiri dari N orang merupakan sebuah minimasi dari sebuah asusmsi dari sebuah distribusi regular dari sekumpulan tanggal lahir. Sedangkan menurut matematikawan lainnya dalam tulisannya menjelaskan kejadian ini ke dalam sebuah tipe pencocokan yang lebih general. Menurut Hocking dan Schwertman dalam sejumlah N individual terdapat probabilitas terbentuknya k pasangan dengan rumus:

$$
\begin{aligned}
P_{N,k} &= P(\text{exactly } k \text{ pairs out of } N \text{ individuals}) \\
&= \frac{N!365!}{(365)^N k! 2^k (N-2k)!(365-N+k)!}.
\end{aligned}
$$

Perhitungan ini merupakan bentuk generalisasi dari Hocking dan Schwertman terhadap kasus kesamaan tanggal lahir pada anak kembar, kembar tiga, kembar empat, dan seterusnya

Perhitungan tersebut pada akhirnya akan mengarahkan perhitungan banyaknya pasangan yang mungkin, bukan sebagai jumlah dari orang yang memliki tanggal lahir yang sama.

III. PERHITUNGAN DAN TEORI

A. Perhitungan kemungkinan

Untuk kasus dimana dari 24 orang tersebut terdapat paling sedikit 2 orang dengan tanggal lahir yang sama. Untuk mempermudah distribusi darii variasi yang akan terbentuk, abaikan adanya kemungkinan dari 24 orang tersebut ada yang kembar dan kemungkinan ada yang memiliki tanggal ulang tahun 29 Febuari. Dengan menganggap semua tanggal memiliki kemungkina yang sama, maka perhitungan mengenai kemungkinan terdapat 2 orang memliki tanggal uang tahun yang sama dan yang tidak memiliki tanggal ulang tahun yang sama dapat dilakukan dengan cara

$$P(A) = 1 - P(A')$$

Dengan;

P(A) = kemungkinan terdapat 2 orang yang memiliki tanggal ulang tahun yang sama

P(A’) = kemungkinan tidak terdapat 2 orang yang memiliki tanggal ulang tahun yang sama.

Apabila kemungkinan P(A) dari 24 orang tersebut sebesar 50% maka jumlah 24 oarng tersebut dapat diambil sebagi contoh .

Ketika kejadian merupakan kejadian yang independen, maka kemungkinan seluruh kejadian terjadi sama dengan kemungkinan salah satu dari kejadian-kejadian tersebut terjadi. Sehingga, dalam kasus 24 orang setiap kejadian adalah kejadian dimana masing masing orang dibandingkan tanggal lahirnya (sesuai dengan kasus tertentu) dengan tanggal lahir 23 orang lainnya keudian dihitung kemnugkinan nya. Dan pada akhirnya kemungkinan terjadinya seluruh kejadian adalah dengan mengalikan setiap kejadian yang terjadi menyangkut kejadian utama tersebut.

Secara penulisan matematik, perhitungan menjadi :

$$P(A') = P(1).P(2).P(3)\ldots.P(23)$$

Untuk kasus pertama , dimana sebelumnya tidak ada orang yang dianalisis.Maka untuk orang pertama yang dianalisis mengalami ketidaksamaan tanggal ulang tahun dengan orang sebelumnya memliki kemungkinan sebesar 100%(365/365) (dengan tidak memperdulikan tahun kabisat).

Untuk kasus kedua, dimana orang kedua dibandingkan dengan orang yang dianalisis sebelumnya, yaitu orang pertama, maka kemungkinan orang tersebut untuk tidak memiliki kesamaan tanggal ulang tahun dengan orang sebelumnya adalah (364/365).Kemungkinan ini didasarkan apabila orang kedua memiliki tanggal lahir yang berbeda dari 364 hari lainnya selain tanggal kelahiran orang pertama.Untuk kasus berikutnya perhitungan kemungkinan dilakukan dengan cara yang sama.

Sehingga kemungkinan terjadinya ketidaksamaan tanggal lahir pada 24 orang yang dipilih secara acak adalah:

P(A’)= 365/365. 364/365. 363/365… .342/365

P(A’)= 0.4616557421

Maka,

P(A)=1-0.4616557421= 0.5383442579(53.834426%)

Proses ini dilakukan dengan memandang kemungkinan terbeut kemungkinan dalam sebuah grup dengan n orang 2 diantaranya memiliki tanggal lahir yang sama

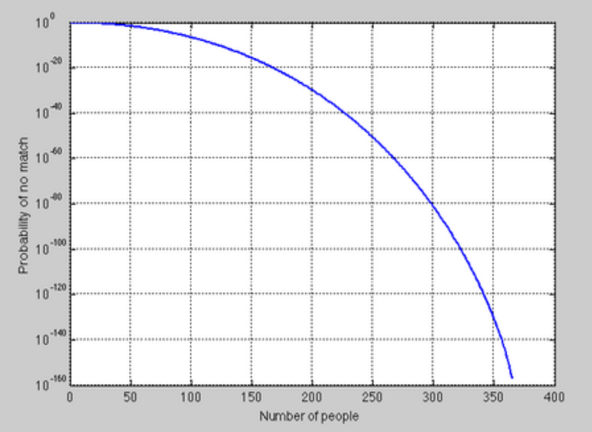
Untuk memempermudah perhitungan, berdasarkan teori pigeonhole:

$$
\begin{aligned}
\bar{p}(n) &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{n-1}{365}\right) \\
&= \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n} \\
&= \frac{365!}{365^n (365-n)!} = \frac{n! \cdot \binom{365}{n}}{365^n}
\end{aligned}
$$

Dengan *p(n)* adalah kemungkinan tidak terdapat pasangan dengan tanggal ulangtahun yang sama. Dan ketika n>365 *p(n)* sama dengan 0. Sama seperti cara mengitung kemungkinan sedikitnya 2 orang yang memiliki tanggal ulang tahun yang sama maka kemungkinan tersebut didapatkan dengan cara:

$$p(n) = 1 - \bar{p}(n).$$

Apabila dibentuk tabel yang nantinya akan menggambarkan grafik kemungkinan, maka grafiknya akan berbentuk :



Dengan Tabel penunjuk grafik:

n	***p*(*n*)**
10	11.7%
20	41.1%
24	53.834%
30	70.6%
50	97.0%
57	99.0%
100	99.99997%
200	99.99999999999999999999999999998%
300	(100 − (6×10^{-80}))%
350	(100 − (3×10^{-129}))%
365	(100 − (1.45×10^{-155}))%
366	100%

B. Aproksimasi

Probabilitas dari kejadian tersebut dapat dicari lagi dengan cara lain dengan hasil yang lebih sederhana dan akurat. Salah satu cara untuk mendapatkan hasil tersebut adalah dengan menggunakan cara eksponensasi sederhana:

$$p(n) \approx 1 - \left(\frac{364}{365}\right)^{C(n,2)} .$$

Dengan angka (364/365) merupakan probabilitas dari dua orang (acak) yang tidak memiliki kesamaan tanggal ulang tahun. Sedangkan kombinasi yang menjadi pangkat dari angka tersebut menandakan jumlah pasangan/ jumlah kejadian yang ada untuk menentukan kemungkinan total kejadian tersebut.

C. Generalisasi (Berdasarkan Ahli)

Dengan menganggap sebagai collision problem:
Diberikan n bilangan integer acak dari discrete uniform distribution dengan range [1,d].Berapa kemungkinan p(n;d) dengan paling sedikit 2 angka sama?(d=365)
Maka,

$$n(p;d) \approx \sqrt{2d \cdot \ln\left(\frac{1}{1-p}\right)}.$$

Dimana dengan konteks ini, birthday problem lebih umum diaplikasikan menurut fungsi hash.

Teori ini digunakan oleh Zoe Schnabel dibawah nama “capture-recapture” statistiv untuk menghitung populasi ikan di danau.

D. Teori yang digunakan untuk pemecahan masalah.

Dalam mencari kemungkinan terjadinya kejadian, terdapat beberapa teori yang mendasari perhitungan dan cara mencari kemungkinan.

a. Pigeonhole Principle
“Apabila n item diletakkan ke dalam m pigeonholes dengan n>m, maka sedikitnya satu pigeonhole harus terdiri dari lebih dari1 item”
Pembentukan ide ini dilakukan oleh Johann Dirichlet ada tahun 1834 dengan judul “Schubfachprinzip”.
Meskipun aplikasi yang paling mudah adalah dengan set terbatas, prinsip in juga digunakan untuk set tidak terbatas yang tidak bisa diletakkan dengan korespondensi satu-satu.Untuk melakukan hal demikian digunakan pernyataan formal bahwa”Fungsi injeksi tidak ada dalam set terbatas dengan kodomainnya lebih kecil daripada domainnya”.
Siegel’s lemma dibentuk dengan prinsip (secara umum) ini.

b. Birtday Attack
Sebuah birthday attack merupakan salah satu tipe dari kriptoanalisi yang mengatur/ merancang perhitungan matematikuntuk pemecahan masalah "Birthday Problem".

c. Fungsi Hash
Fungsi Hash merupakan algoritma yang memetakan set data besar, yang juga disebu kunci, ke set data yang lebih kecil.
Contohnya untuk bilangan integer tunggal dapat berfungsi sebagai indeks daripada sebuah array. Nilai yang dikembalikan oleh fungsi hash disebut nilai hash, kode-kode hash, jumlah hash, atau hash-hash sederhana.
Fungsi hash terhubung (dan sering dibingungkan dengan) checksums, sek digit-digit, fingerprints(sidik jari), pengacakan fungsi-fungsi,kode-kode pengoreksi eror dan fungsi hash kriptografi. Meskipun bergitu, konsep ini "bertumpukan" sampai batas tertentu,dimana masing-masing kondep memiliki kegunaan tertentu dan syarat-syarat serta dirancang dan dioptimasi secara berbeda.

IV. CONTOH KASUS BIRTHDAY PROBLEM

Pada kejadian jumlah pasang karyawan yang memiliki tanggal lahir yang sama dalam suatu perusahaan di Minnesotta (berdasarkan pengamatan dari seorang *William J.Poley*)

William menemukan suatu kejadian unik pada suatu surat kabar yang menyatakan bahwa mereka menemukan suatu kejanggalan dari sebuah perusahaan yang memiliki sejumlah pasang karyawan yang memiliki tanggal lahir yang sama. Surat kabar tersebut menemukan pada tahun 1998, perusahaan tersebut memiliki 37 karyawan dan memiliki 4 buah pasang karyawan dengan tanggal lahir yang sama, dan pada tahun berikutnya, dengan 36 karyawan, perusahaan tersebut memiliki 5 buah pasang karyawan dengan tanggal lahir yang sama.

Dengan melihat beberapa penemuan dari matematikawan yang ada terdapat teori yang mungkin dapat memenuhi kecurigaan dari surat kabar tersebut terkait kejadian yang terjadi pada perusahaan tersebut. Tapi , dalam teori yang mungkin dapat menjawab kecurigaan tersebut, teori tersebut dibentuk dengan asusmsi bahwa karyawan tersebut merupakan saudara kembar, baik kembar dua, tiga dan seterusnya. Sedangkan pada kenyataannya, karyawan yang bekerja dalam perusahaan tersebut merupakan karyawan yang selalu berganti-ganti dari tahun ketahun.

Untuk menjawab hal tersebut, diterapkan sebuah teori yang mempertimbangkan sirkulasi dari karyawan yang ada di perusahaan tersebut. Teori tersebut disebut teori rantai Markov.

Rantai Markov ini mengandung paham bahwa ketika satu orang keluar dari perusahaan tersebut, terdapat satu orang yang masuk ke dalam perusahaan tersbut.

Untuk jumlah kecil N (N=4)

- Terdapat 5 state yang mungkin : tidak ada yang cocok, satu buah pasang, satu buah triplet, satu buah quartet dan 2 buah pasang.
- Terdapat asumsi bahwa setiap orang memiliki probabilitas yang sama untuk meninggalkan perusahaan dan orang baru yang masuk perusahaan,yang merupakan penarikan acak dari populasi dengan distribusi regular dari tanggal ulang tahun.
- Terdapat asumsi bahwa perusahaan akan menerima karyawan baru untuk mengganti karyawan yang telah meninggalkan perusahaan.

Probabilitas ketika seorang karyawan keluar dan seorang karyawan baru masuk adalah $T_{i,j}$. Ketika i = tidak cocok dan j= tidak cocok , maka terdapat 4 cara unutk memilih siapa yang keluar dan 365 cara untuk memilih karyawan baru. Maka penyebut dari pencarian probabilitas $T_{i,j}$ adalah 1460 dan pembilang dari perhitungan probabilitas adalah banyak cara dari seseorang bisa meniggalkan perusahaan dan yang lainnya dapat masuk ke perusahaan, tapi masih tidak memberikan kecocokan dalam tanggal lahir. Dan probailitas ketika tidak ada kecocokan tanggal lahir dari satu orang yang keluar dari perusahaan dan satu orang yang masuk ke dalam perusahaan adalah 1448/1460. Ketika seluruh kemungkinan didapatkan, kemungkinan tersebut kemudian dituliskan dalam sebuah matriks seperti berikut:

$$
\begin{array}{l c}
 & \begin{array}{ccccc} \text{no match} & \text{1 pair} & \text{1 triple} & \text{1 quad} & \text{2 pair} \end{array} \\
\begin{array}{l} \text{no matches} \\ \text{1 pair} \\ \text{1 triple} \\ \text{1 quadruple} \\ \text{2 pairs} \end{array} &
\begin{bmatrix}
\frac{4}{4}\frac{362}{365} & \frac{2}{4}\frac{362}{365} & 0 & 0 & 0 \\
\frac{4}{4}\frac{3}{365} & \frac{2}{4}\frac{2}{365}+\frac{2}{4}\frac{364}{365} & \frac{3}{4}\frac{363}{365} & 0 & \frac{4}{4}\frac{363}{365} \\
0 & \frac{2}{4}\frac{1}{365} & \frac{3}{4}\frac{1}{365}+\frac{1}{4}\frac{364}{365} & \frac{4}{4}\frac{364}{365} & \frac{4}{4}\frac{1}{365} \\
0 & 0 & \frac{1}{4}\frac{1}{365} & \frac{4}{4}\frac{1}{365} & 0 \\
0 & \frac{2}{4}\frac{1}{365} & \frac{3}{4}\frac{1}{365} & 0 & \frac{4}{4}\frac{1}{365}
\end{bmatrix}
\end{array}
$$

Pada matriks terdapat sebuah vektor yang menggambarkan probablitas ketika berada di salah satu state pada waktu t. Seperti yang disebutkan dalam pejelasan terkait apa yang terjadi dalam suatu waktu t, vektor waktu berikutnya adalah :

$$s_{t+1} = Ts_t.$$

Untuk membuktikan berlakunya teori ini, dilakukan sebuah pengamatan terkait matriks yang ada dalam waktu tertentu dan kejadian tertentu (ketika seorag karyawan keluar dan seorang karyawan baru masuk). Dalam pegamatannya terdapat sejumlah definisi dari rantai Markov yang dapat digunakan untuk menyimpulkan probabilitas yang terjadi.:

Definisi 1 : Sebuah rantai Markov akan menjadi ketika setiap state terjadi kembali (recurrent). Hal ini kemudian akan mengembalikan proses ke dalam state yang diketahui dengan probabilitas sama dengan 1.

Definisi 2 : Sebuah rantai Markov dikatakan regular jika terdapat seebuah integer positif l dan T^l hanya mengandung elemen positef saja.

Definisi 3 : Sebuah distribusi invariant s , adalah sebuah vektor yang memenuhi

$$s = Ts.$$

Sebuah distribusi invariant sering kali dikaitkan sebagai distribusi dari state yang tetap.

Untuk N=4, perhitungan sederhana yang menyimpulkan bahwa terdapat 3 cara untuk bergerak dari sebuah state peluang yang satu ke state peluang lainnya. “+” merupakan symbol yang menunjukkan elemen yang lbernilai lebih besar dari nol.

$$\begin{bmatrix} + & + & 0 & 0 & 0 \\ + & + & + & 0 & + \\ 0 & + & + & + & + \\ 0 & 0 & + & + & 0 \\ 0 & + & + & 0 & + \end{bmatrix}^3 = \begin{bmatrix} + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{bmatrix}$$

Dan eigenvector yang menggambarkan permaslahan ini adalah

$$s = \begin{bmatrix} \frac{47831784}{48627125} \\ \frac{792792}{48627125} \\ \frac{1456}{48627125} \\ \frac{1}{48627125} \\ \frac{1092}{48627125} \end{bmatrix}.$$

Probabilitas ini sama dengan probabilitas yang dihasilkan oleh percobaan terpisah, memilih 4 orang dari sebuah populasi yang memiliki distribusi tanggal lahir yang regular.

Rumus dari Hocking dan Schwertman untuk k_3 tiga pasang dan k_2 pasang dari N menghasilkan elemen ketiga dari vektor

$$P_{k_3,k_2} = \left[\frac{N!}{(3!)^{k_2}(2!)^{k_2}(365)^N}\right]\left[\frac{365!}{k_3!k_2!(N-3k_3-2k_2)!(365-n+2k_3+k_2)!}\right]$$

$$P_{1,0} = \left[\frac{4!}{(3!)(365)^4}\right]\left[\frac{365!}{(365-4+2)!}\right] = \frac{1456}{48627125}.$$

Probabilitas yang keempat adalah probabilitas dimana semua orang memiliki ulang tahun yang sama, yaitu

$$\frac{365}{365^4} = \frac{1}{48627125}.$$

Pada kenyataannya, ketika dikaitkan dengan asumsi yang dilakuakn di awal terdapat ketidak cocokan asumsi. Terdapat kejadian-kejadian yang menyebabkan munculnya peningkatan jumlah pasangan dengan tanggal lahir yang sama dikarenakan terdapat siklus penerimaan dan pemecatan karyawan secara drastis. Hal ini menyebabkan terdapat kecurigaan dari penulis artikel pada surat kabar terkait kejadian ini.

KESIMPULAN

Persoalan tanggal lahir ini merupakan salah satu persoalan yang dapat diselesaikan dengan berbagai cara. Salah satu cara yang dibahas dalam contoh kasus adalah dengan rantai Markov. Dengan mempertimbangkan beberapa kemungkinan yang ada dalam ruang sampel , kita dapat menentukan probabilitas yang mendekati tepat dengan kedaan sesungguhnya. Prinsip ini lebih dikenal dengan istilah “revolving door”, yang dimana dengan menggunakan prinsip ini, pembelajaran terkait probabilitas dalam perosoalan ini dapat lebih dimengerti dan dipahami.

REFERENSI

[1] Coincidences : The Truth is Out There
http://www.rsscse-edu.org.uk/tsj/wp-content/uploads/2011/03/matthews.pdf
Waktu akses :17 Desember 2012; pukul 9.18 WIB

[2] A Revolving Door Birthday Problem
http://williampolley.com/webpapers/birthday.pdf
Waktu akses : 17 Desember 2012;pukul 9.20 WIB

[3] Birthday Problem- Salem Press
http://salempress.com/store/pdfs/birthday.pdf
Waktu akses : 17 Desember 2012;pukul 9.20 WIB

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 17 Desember 2012

Aurelia H B Matondang-13510023