

# Pembuktian Hukum Zipf terhadap Karakter Alfabet dari Data Nama Mahasiswa ITB

**Azka Hanif Imtiyaz - 13514086**

*Program Studi Teknik Informatika*

*Sekolah Teknik Elektro dan Informatika*

*Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia*

*13514086@std.stei.itb.ac.id*

**Abstrak**—Hukum Zipf merupakan hukum matematika sederhana yang mendasari frekuensi distribusi dari kata-kata yang dibuat oleh manusia. Pada makalah ini akan dibuktikan bahwa hukum Zipf tidak hanya mendasar frekuensi dari kata-kata, tetapi juga karakter alfabetis.

**Kata kunci**—Alfabet, Hukum Zipf, Nama, Statistika.

## I. PENDAHULUAN

George Kingsley Zipf merupakan ahli bahasa dan filolog dari Amerika. Beliau mempelajari kejadian statistik dari berbagai bahasa di dunia. Beliau adalah yang mengemukakan hukum Zipf yang hingga saat ini masih kontroversial dalam pembuktiannya secara umum.

Zipf mengemukakan bahwa semakin sering suatu kata yang digunakan pada suatu bahasa tertentu akan menyebabkan kata tersebut semakin sering untuk digunakan kembali, namun semakin jarang suatu kata yang digunakan pada suatu bahasa tertentu akan menyebabkan kata tersebut semakin jarang untuk digunakan kembali. Sehingga menurut hipotesis tersebut Zipf dapat membuat hukum Zipf.

Hukum Zipf pada awalnya hanya berlaku pada sastra saja, namun seiring berjalannya waktu, ditemukan bahwa hukum Zipf ini berlaku pula pada frekuensi kejadian lainnya yang melibatkan manusia dan interaksinya. Hal ini menyebabkan kontroversi, beberapa ilmuwan mengatakan bahwa hal ini hanyalah keberuntungan statistik dan dapat disangkal dengan pembuktian pada ketidak benarannya. Namun, sejauh ini hukum Zipf masih belum dapat dibuktikan ketidak benarannya, dan malah banyak hukum turunan yang didapat dari hukum Zipf tersebut.

Hukum Zipf membuktikan bahwa interaksi manusia, tidak peduli seacak apapun, akan mengikuti suatu pola, yang banyak makin banyak, begitu pula yang sedikit. Hal ini mendasari suatu prinsip yang baru yaitu prinsip Pareto sebagai turunan dari hukum Zipf yang ternyata selama ini sering diaplikasikan pada bidang manapun dalam kehidupan. Prinsip ini mengatakan bahwa terdapat suatu rasio penyebab 20:80 dan akibat 80:20. Sebagai contoh, 20% pelanggan dari suatu perusahaan akan menghasilkan 80% dari total pemasukan perusahaan tersebut. Contoh lainnya, 20% pelanggan dari suatu perusahaan akan

mengakibatkan permasalahan sebesar 80% pada perusahaan tersebut. Hal ini menyebabkan fokus untuk menyelesaikan suatu permasalahan dapat diarahkan kepada 20% tersebut, maka permasalahan akan selesai.

Pada makalah ini akan diaplikasikan dan membuktikan kebenaran hukum Zipf pada karakter alfabet yang terdapat pada nama dan nim mahasiswa ITB.

## II. TINJAUAN PUSTAKA

### 1. Induksi Matematik

Induksi matematik adalah metode pembuktian yang dapat digunakan untuk menentukan kebenaran dari suatu pernyataan yang diberikan dalam bentuk bilangan bulat.

Di dalam induksi matematik terdapat dua langkah dalam pembuktiannya, yaitu basis dan induksi. Basis merupakan suatu pernyataan dasar yang merupakan batas bawah dari limit suatu bilangan bulat dari pernyataan yang ingin dibuktikan yang bernilai benar atau sah, basis ini kemudian dilanjutkan ke bilangan bulat selanjutnya dengan induksi. Induksi merupakan langkah selanjutnya dari basis, setelah basis berhasil dibuktikan, maka pembuktian akan dilakukan dengan cara membuktikan langkah-langkah selanjutnya yang dapat membuktikan bahwa langkah-langkah setelah langkah yang dibuktikan bernilai benar atau sah. Jika basis dan Induksi bernilai benar atau sah, maka pernyataan yang ingin dibuktikan telah terbukti.

Dalam pembuktian suatu pernyataan, terkadang metode induksi matematik memerlukan asumsi bahwa pernyataan yang ingin dibuktikan adalah benar atau sah. Asumsi ini disebut hipotesis induksi.

Karena langkah induksi yang secara tidak langsung membuktikan pernyataan pada suatu  $n$ , yaitu kasus setelah induksi tersebut dibuktikan, maka metode induksi matematik dapat diibaratkan seperti efek domino.



**Gambar 2-1** Induksi matematik yang diibaratkan seperti domino.

Sumber:

<http://www.thedailysheep.com/wp-content/uploads/2014/08/seo-domino.jpg>

## 2. Hukum Zipf

Hukum Zipf menyatakan terdapat korelasi antara frekuensi muncul suatu kata terhadap urutan peringkat setelah diurutkan terhadap karakter lain sesuai frekuensi. Suatu peringkat  $r$ , dimana  $r$  merupakan urutan frekuensi kemunculan kata, memiliki frekuensi  $f(r)$  yang mengikuti aturan sebagai berikut.

$$f(r) \sim \frac{1}{r^\alpha} \quad (1)$$

Dengan  $\alpha \approx 1$  (Zipf, 1936, 1949), dengan suatu batas kesalahan tertentu. Aturan ini kemudian dikembangkan agar mendapatkan nilai pasti dari suatu  $r$ . Maka apabila  $n$  adalah frekuensi kemunculan dari peringkat pertama, aturannya adalah sebagai berikut.

$$f(r) = \frac{1}{r^\alpha} n \quad (2)$$

Dengan dikembangkannya menjadi bentuk di atas, maka batas kesalahan yang didapat dari hasil persamaan akan lebih besar, namun tetap mengikuti rasio tersebut. Dengan batasan  $r \geq 1, n = f(1), \alpha = 1$ .

## 3. Data yang Digunakan

Data yang digunakan merupakan nama dari seluruh mahasiswa IF ITB angkatan 2014 dan juga nama dari seluruh mahasiswa ITB angkatan 2011 hingga 2014.

Data yang akan diambil adalah frekuensi karakter alfabet (a hingga z) yang ada pada nama-nama mahasiswa tersebut. Frekuensi karakter ini kemudian akan diurutkan membesar dan diberikan peringkat masing-masing.

## III. PEMBUKTIAN

### 1. Induksi Matematik

Basis- Frekuensi dari peringkat pertama adalah:

$$f(1) = \frac{1}{1} f(1) \quad (3)$$

Sehingga basis terbukti.

Induksi- Frekuensi dari suatu peringkat  $n$  adalah:

$$f(n) = \frac{1}{n} f(1) \quad (4)$$

Bernilai benar, karena  $n \geq 1$ . Maka  $n + 1$  akan bernilai benar pula, sehingga pembuktian secara matematis terbukti.

Namun pembuktian secara matematik saja tidak cukup karena jelas hukum Zipf berlaku terhadap suatu data tertentu, sehingga pembuktian dari hukum Zipf akan dilakukan terhadap data nama dan nim mahasiswa IF ITB angkatan 2014, kemudian mahasiswa ITB angkatan 2011 hingga 2014.

## 2. Nama Mahasiswa IF ITB Angkatan 2014

Peringkat	Alfabet	Frekuensi
1	a	374
2	i	166
3	n	133
4	r	132
5	s	94
6	h	90
7	m	90
8	t	89
9	u	89
10	d	79
11	e	67
12	o	65
13	l	56
14	k	40
15	y	38
16	f	32
17	j	25
18	p	25
19	g	21
20	v	21
21	c	20
22	w	18
23	z	17
24	b	12
25	q	2
26	x	0

Tabel 3-1 Karakter yang ada pada nama mahasiswa IF ITB angkatan 2014

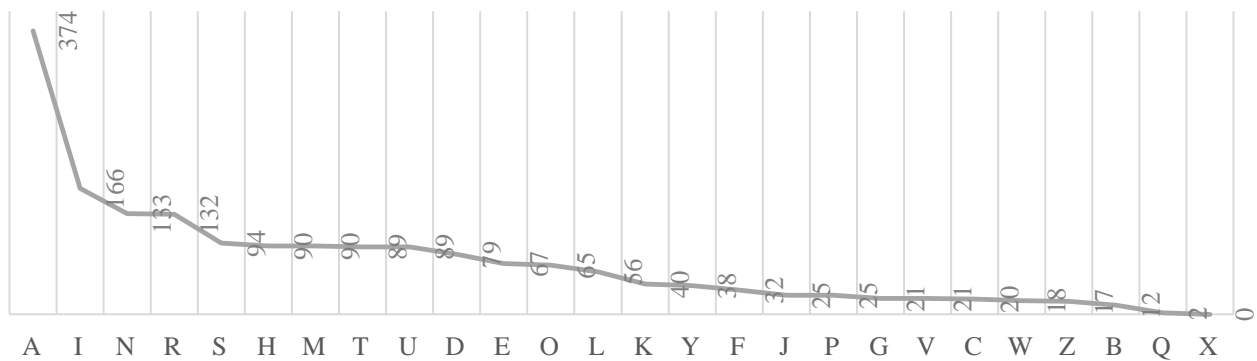
Basis- Dari tabel didapatkan bahwa  $f(1) = 374$ .

Induksi-  $f(2)$  seharusnya bernilai:

$$f(2) = \frac{1}{2} f(1) = \frac{1}{2} 374 = 187 \quad (5)$$

Namun  $f(2)$  ternyata memiliki nilai 166. Hal ini wajar karena akan terdapat suatu batas kesalahan tertentu, namun rasio dari masing-masing frekuensi terhadap frekuensi pertama seharusnya tetap menurut hukum Zipf, berikut adalah grafik yang dibuat berdasarkan tabel di atas.

### FREKUENSI ALFABET YANG ADA PADA NAMA MAHASISWA IF 2014



Grafik 3-1 Grafik peringkat karakter terhadap frekuensi kemunculan karakter tersebut.

Grafik ini kemudian akan dibandingkan dengan grafik dari alfabet dari nama mahasiswa ITB angkatan 2011 hingga 2014.

#### 3. Nama Mahasiswa ITB Angkatan 2011 Hingga 2014

1	a	47168
2	i	26064
3	n	18110
4	r	16770
5	d	10603
6	h	10452
7	u	10248
8	m	10070
9	s	10001
10	t	9749
11	l	7480
12	e	7454
13	o	6090
14	y	6068

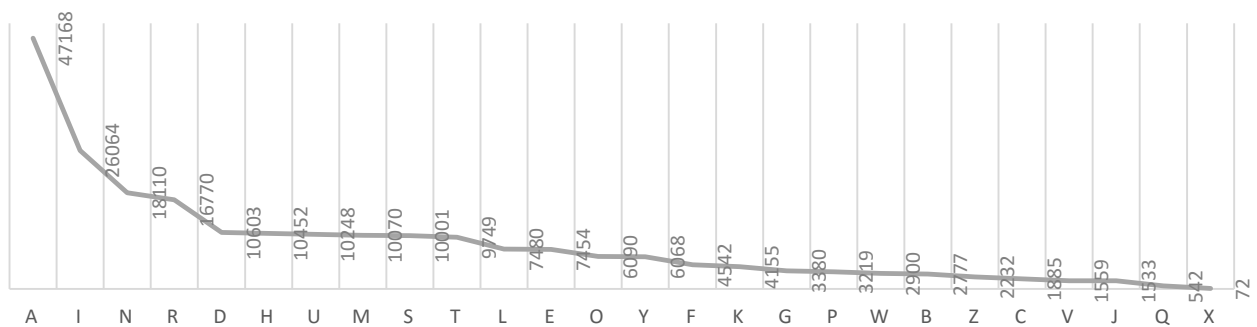
15	f	4542
16	k	4155
17	g	3380
18	p	3219
19	w	2900
20	b	2777
21	z	2232
22	c	1885
23	v	1559
24	j	1533
25	q	542
26	x	72

Tabel 3-2 Karakter yang ada pada nama mahasiswa ITB angkatan 2011 hingga 2014

Basis- Dari tabel didapatkan bahwa  $f(1) = 47168$ .  
Induksi-  $f(2)$  seharusnya bernilai:

$$f(2) = \frac{1}{2}f(1) = \frac{1}{2}47168 = 23584 \quad (6)$$

### FREKUENSI ALFABET YANG ADA PADA NAMA MAHASISWA ITB



Grafik 3-2 Grafik peringkat karakter terhadap frekuensi kemunculan karakter tersebut.

Namun  $f(2)$  ternyata memiliki nilai 26064. Hal ini wajar karena akan terdapat suatu batas kesalahan tertentu, namun rasio dari masing-masing frekuensi terhadap frekuensi pertama seharusnya tetap menurut hukum Zipf, Grafik 3-2 dibuat berdasarkan data pada tabel 3-2.

#### IV. PENGOLAHAN DATA

Berdasarkan data, tabel, dan grafik yang telah dibuat sebelumnya, maka akan dibuat grafik dalam bentuk logaritma berbasis bilangan euler untuk melihat rasio dari masing-masing peringkat.



Grafik 4-1 Grafik rasio yang dibuat dari Tabel 3-1



Grafik 4-2 Grafik rasio yang dibuat dari Tabel 3-2

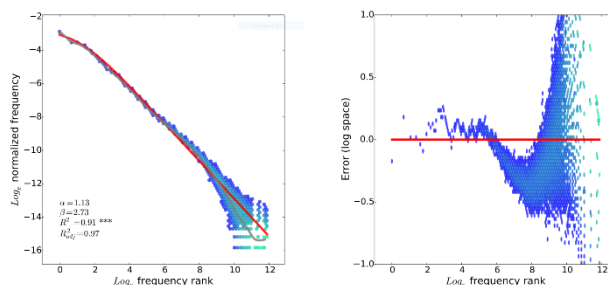
Ternyata rasio dari masing masing grafik serupa dengan batas kesalahan tertentu. Maka dapat pembuktian secara induksi matematik menyatakan bahwa hukum Zipf terhadap karakter alfabet dari data nama mahasiswa ITB adalah sah dengan batas kesalahan tertentu.

Walaupun tidak semua peringkat dari karakter mempunyai rasio yang konsisten, tetapi dari perbandingan dua data yang telah diambil dapat dilihat bahwa semakin banyak data yang diambil dan diolah, maka semakin akurat hukum Zipf tersebut. Hal ini terlihat dari rasio peringkat dari “q” dan “x” pada data nama mahasiswa IF ITB angkatan 2014 menurun drastis, sedangkan peringkat dari

“q” dan “x” pada data nama mahasiswa ITB angkatan 2011 hingga 2014 lebih konsisten

#### V. KESIMPULAN

Hukum Zipf terhadap karakter alfabet dari data nama mahasiswa ITB telah terbukti sah dengan batas kesalahan tertentu. Dari sini terlihat bahwa hukum Zipf ternyata tidak hanya melandasi kata-kata, tetapi juga karakter alfabetik pada suatu nama. Walaupun banyak sekali kontroversi mengenai hukum Zipf tersebut, tetapi sejauh ini hukum Zipf selalu terbukti benar dengan batas kesalahan tertentu. Berikut adalah grafik dari rasio hukum Zipf secara umum



beserta batas kesalahannya.

Gambar 5-1 Grafik dari rasio hukum Zipf secara umum beserta batas kesalahannya

Sumber:

Zipf 's word frequency law in natural language: a critical review and future directions (Steven T. Piantadosi, 2015)

Hal ini sangat menarik, sebab manusia dan interaksinya seperti tinggi manusia rata-rata, banyaknya mobil manusia yang dikendarai secara umum, hingga frekuensi aktivitas yang dilakukan manusia sehari hari, yang seharusnya tidak mempunyai landasan tertentu dan sangat kompleks untuk digeneralisasikan, ternyata dapat diberikan landasan seperti hukum Zipf.

Bahkan suatu kata yang dibentuk secara acak dari seorang manusia yang mengetik secara acak pada suatu papan ketik, akan mengikuti hukum Zipf. Hal ini membuktikan bahwa hukum Zipf dapat diaplikasikan ke banyak hal, selama hal tersebut melibatkan manusia.

Banyak teori yang membenarkan hukum Zipf, namun teori yang paling diterima adalah teori “*path of least effort*”. Teori ini menyebutkan bahwa manusia akan cenderung melakukan sesuatu hal yang sama apabila hal tersebut sudah sangat sering. “*The rich is richer, the big is bigger, the popular is popular-er. It’s just basic math.*” – Michael Vsauc.

#### REFERENSI

- [1] Steven T. Piantadosi, *Zipf 's word frequency law in natural language: a critical review and future directions.* (2 Juni 2015).
- [2] M. E. J. Newman, *Power laws, Pareto distributions and Zipf's law.* Department of Physics and Center for the Study of Complex

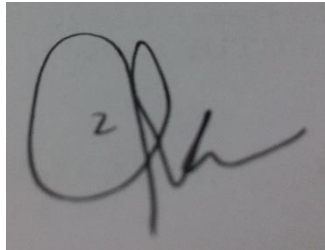
Systems, University of Michigan, Ann Arbor, MI 48109. U.S.A. (29 Mei 2006)

- [3] *Who's afraid of George Kingsley Zipf? Or: Do children and chimps have language?.* Significance, (Desember 2013), hal 29-34.
- [4] ([http://io9.com/the-mysterious-law-that-governs-the-size-of-your-city-1479244159?utm\\_expid=66866090-48.Ej9760cOTJCPS\\_Bq4mjoww.0](http://io9.com/the-mysterious-law-that-governs-the-size-of-your-city-1479244159?utm_expid=66866090-48.Ej9760cOTJCPS_Bq4mjoww.0) diakses pada 10 Desember 2015)
- [5] (<https://plus.maths.org/content/os/latestnews/may-aug08/food/index> diakses pada 9 Desember 2015)
- [6] (<http://judson.blogs.nytimes.com/2009/05/19/math-and-the-city/?em> diakses pada 10 Desember 2015)
- [7] (<https://plus.maths.org/content/mystery-zipf?src=aop> diakses pada 10 Desember 2015)
- [8] (<http://wugology.com/zipfs-law/> diakses pada 10 Desember 2015)

## 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, 10 Desember 2015

A photograph of a handwritten signature in black ink on a light-colored surface. The signature is stylized and appears to be 'Azka Hanif Imtiyaz'.

Azka Hanif Imtiyaz - 13514086