

# Klasifikasi Multi-Label pada Hadis Bukhari dalam Terjemahan Bahasa Indonesia Menggunakan Information Gain dan Backpropagation Neural Network

---

Muhammad Yuslan Abubakar  
1301141071

Prof. Dr. Adiwijaya, S.Si., M.Si  
Said Al Faraby, S.T., M.Sc



# Pendahuluan

---

- Apa itu hadis?
- Pengelompokan hadis
- Masalah pada pengelompokan hadis
- Data hadis
- Metode penyelesaian masalah



# Studi Terkait

---

## 1. Hadis

- Kawther A. Aldhlan – *Decision Tree*
- Al-Kabi – TFIDF
- Eliza J – TFIDF dan ANN
- Andina K – SVM

## 2. Multi-Label

- Min-Ling Z – BPMLL
- J. Huang – *Joint Feature Selection and Classification with Fisher Discriminant-Based*
- H. Liu – Regresi PLS dan  $l_1$ -norm



# Sistem yang Dibangun

- Representasi Data Multi-Label

Data	Anjuran	Larangan	Informasi
Janganlah kalian berdusta terhadapku (atas namaku), karena barangsiapa berdusta atas namaku dia akan masuk neraka.	0	1	1
Kami pernah shalat Maghrib bersama Nabi ketika matahari sudah tenggelam tidak terlihat.	0	0	1
Apabila seorang dari kalian memperbaiki kelslamannya maka dari setiap kebaikan akan ditulis baginya sepuluh (kebaikan) yang serupa hingga tujuh ratus tingkatan, dan setiap satu kejelekan yang dikerjakan akan ditulis satu kejelekan saja yang serupa dengannya.	1	0	1

# Sistem yang Dibangun (cont.)

- Representasi Data Multi-Label

Data	Kelas Kitab
Janganlah kalian berdusta terhadapku (atas namaku), karena barangsiapa berdusta atas namaku dia akan masuk neraka.	0 1 0 0 0
Kami pernah shalat Maghrib bersama Nabi ketika matahari sudah tenggelam tidak terlihat.	0 0 0 0 1
Apabila seorang dari kalian memperbaiki keIslamannya maka dari setiap kebaikan akan ditulis baginya sepuluh (kebaikan) yang serupa hingga tujuh ratus tingkatan, dan setiap satu kejelekan yang dikerjakan akan ditulis satu kejelekan saja yang serupa dengannya.	1 0 0 0 0

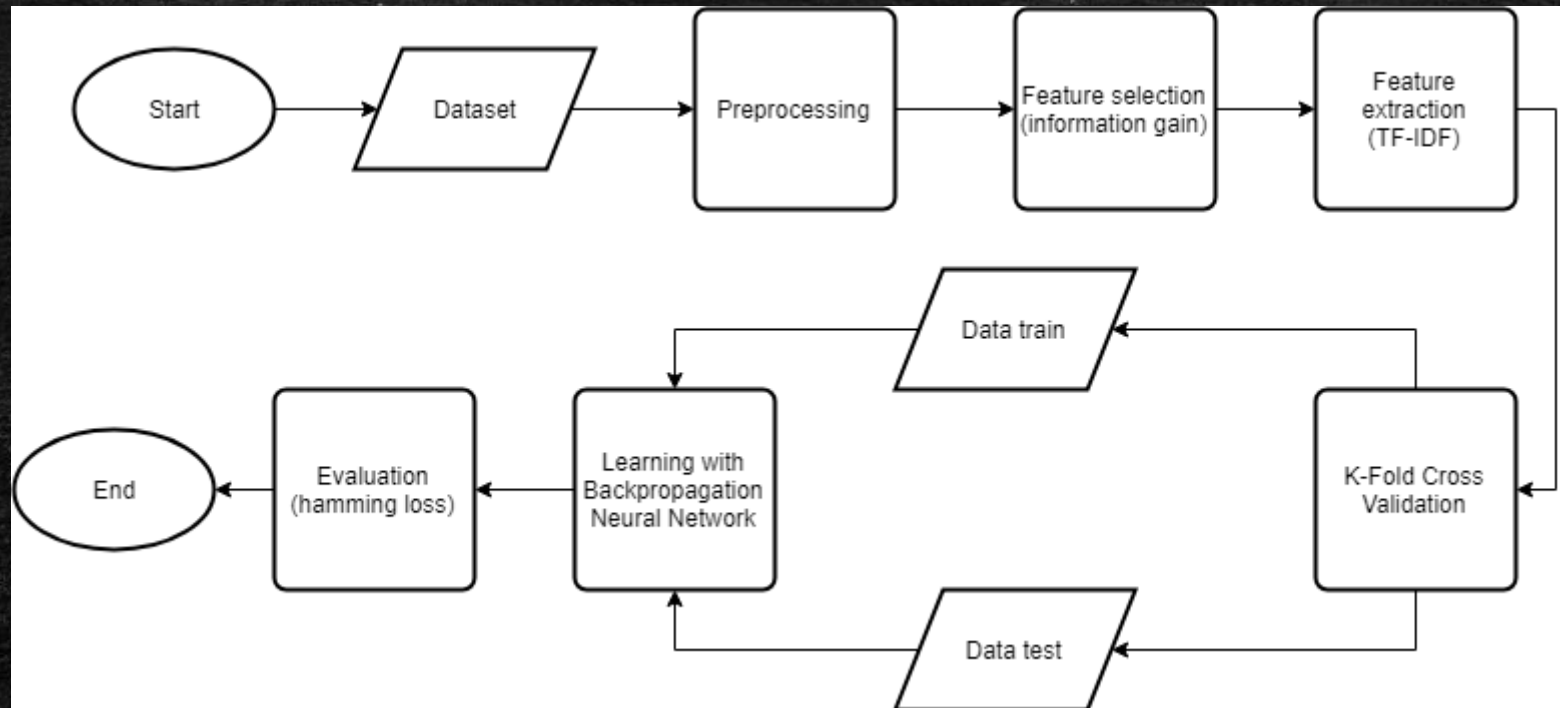
## Data Kelas Kitab

1 0 0 0 0	👉	Iman
0 1 0 0 0	👉	Ilmu
0 0 1 0 0	👉	Wudhu
0 0 0 1 0	👉	Shalat
0 0 0 0 1	👉	Waktu Shalat



# Sistem yang Dibangun (cont.)

- Flowchart Sistem

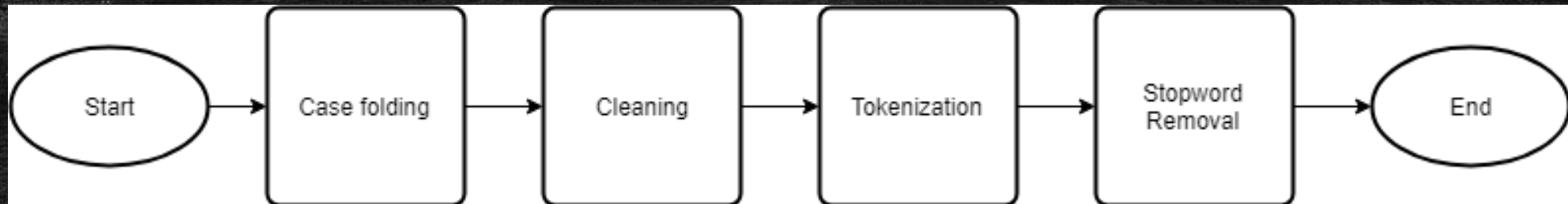




# Sistem yang Dibangun (cont.)

---

- Preprocessing





## Sistem yang Dibangun (cont.)

---

- Feature Selection (Information Gain)

$$I(C, A) = H(C) - H(C|A)$$



$$H(C) = - \sum_{c \in C} p(C) \log p(C)$$

$$H(C|A) = - \sum_{c \in C} p(C|A) \log p(C|A)$$

*dimana :*

$H(C) = - \sum_{c \in C} p(C) \log p(C)$  = merupakan entropi dari kelas

$H(C|A) = - \sum_{c \in C} p(C|A) \log p(C|A)$  = merupakan kondisi entropi dari tiap kelas terhadap atribut/kata



## Sistem yang Dibangun (cont.)

---

- Feature Extraction (TF-IDF)

$$w_{ij} = tf \times idf = tf_{ij} \times \log \frac{D}{df_i}$$

dimana :

$w_{ij}$  = bobot kata  $t_j$  terhadap dokumen  $d_i$

$tf_{ij}$  = jumlah kemunculan kata  $t_j$  dalam dokumen  $d_i$

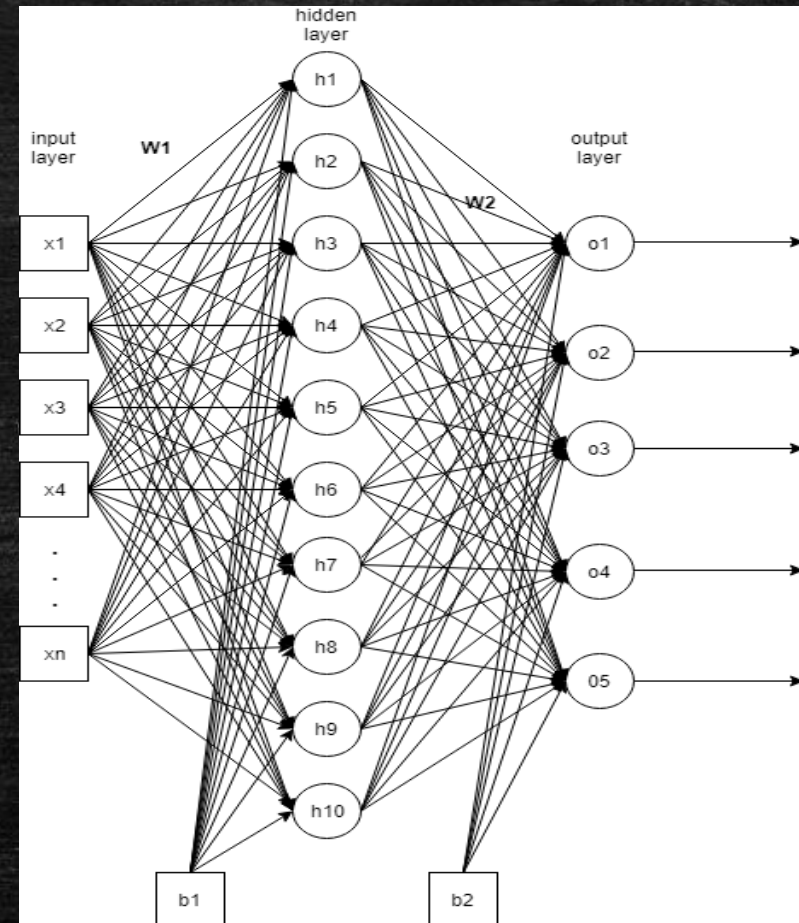
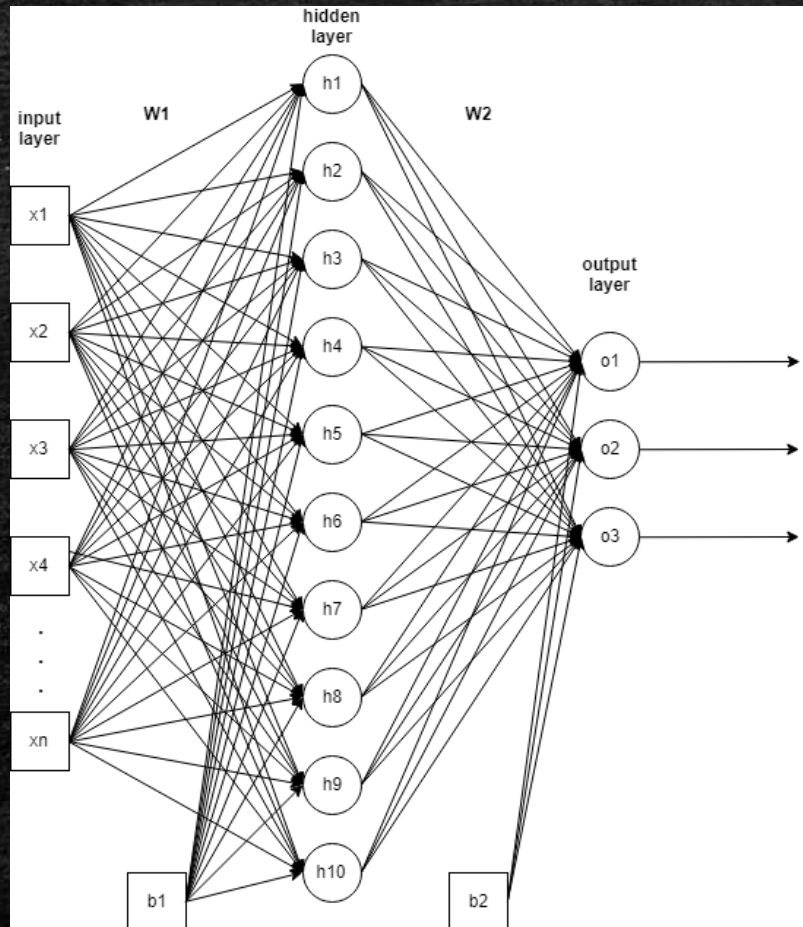
$D$  = jumlah dokumen

$df_i$  = jumlah kemunculan kata dalam dokumen  $D$



# Sistem yang Dibangun (cont.)

- Classifier (Backpropagation Neural Network)





## Sistem yang Dibangun (cont.)

---

- Metriks Pengukuran (Hamming Loss)

$$hLoss(h) = \frac{1}{p} \sum_{i=1}^p \frac{1}{Q} |h(x_i) \Delta Y_i|$$

dimana :

$p$  = banyaknya data yang diklasifikasikan

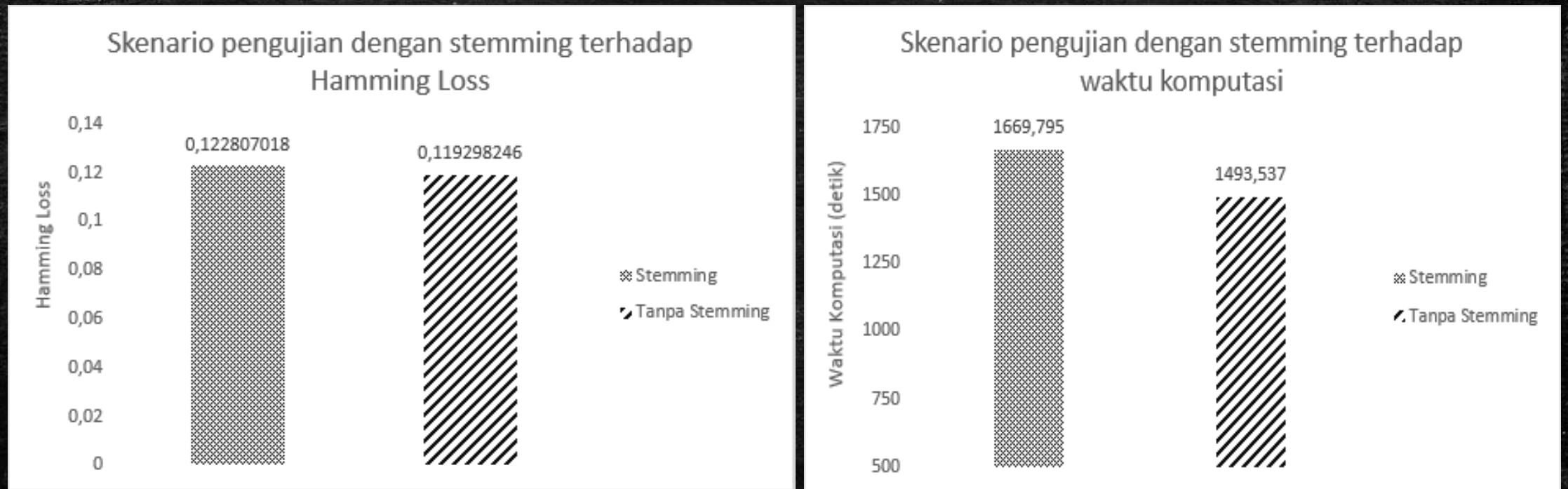
$Q$  = jumlah label kelas

$|h(x_i) \Delta Y_i|$  = banyaknya kesalahan klasifikasi yang terjadi



# Evaluasi

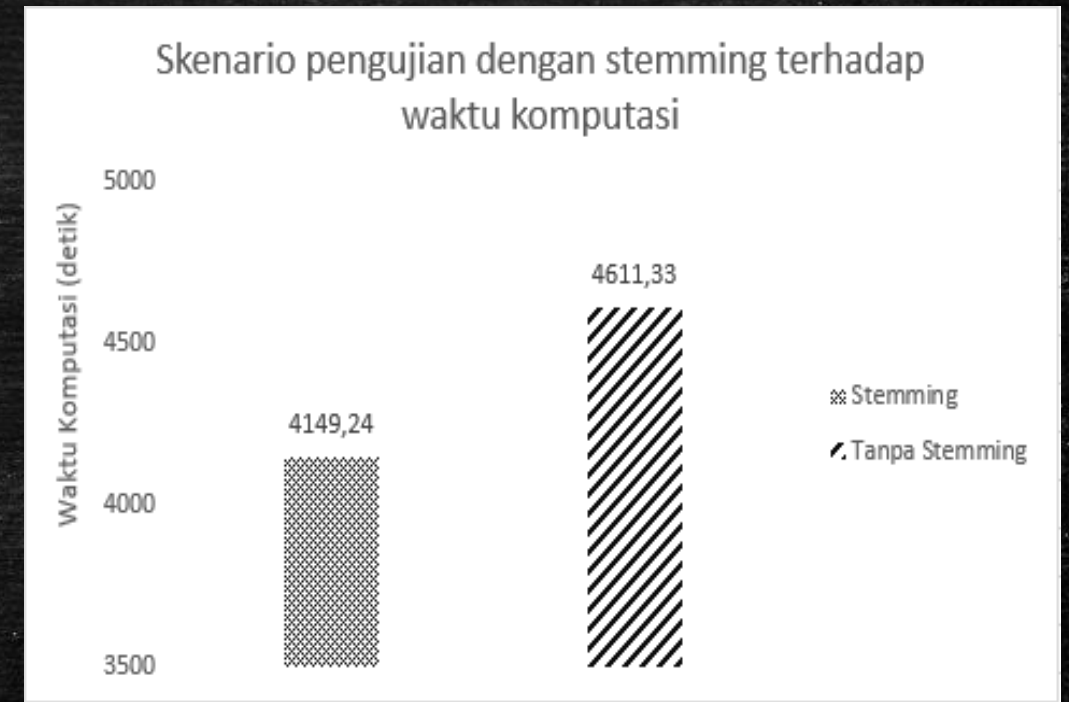
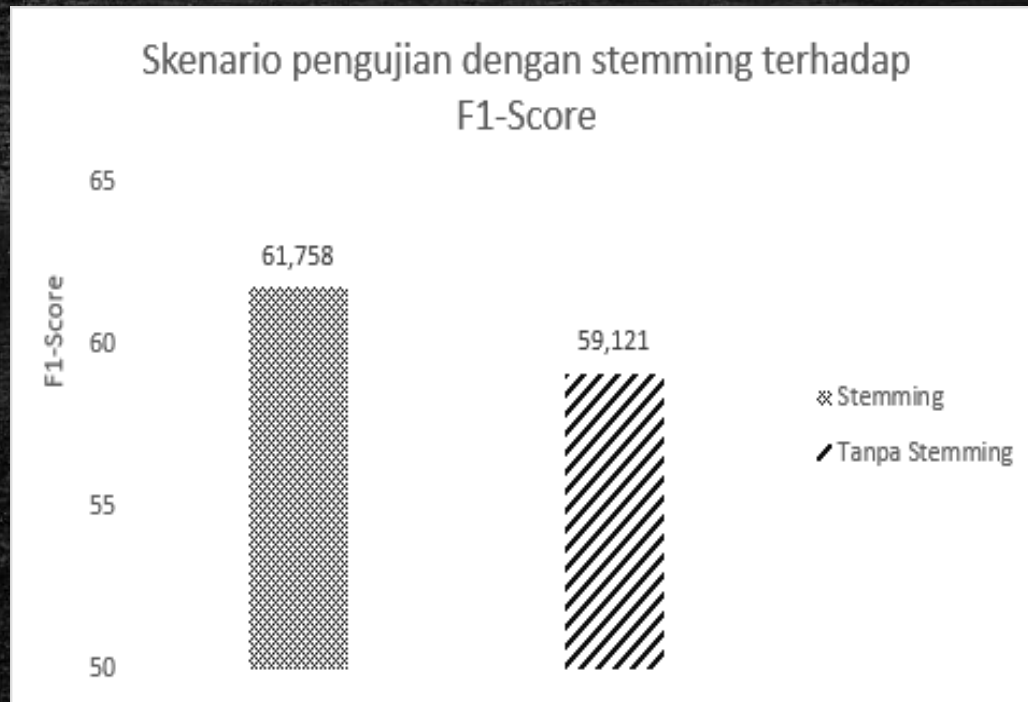
- Skenario Pengujian dengan Proses Stemming (Multi-Label)





# Evaluasi (cont.)

- Skenario Pengujian dengan Proses Stemming (Single Label)

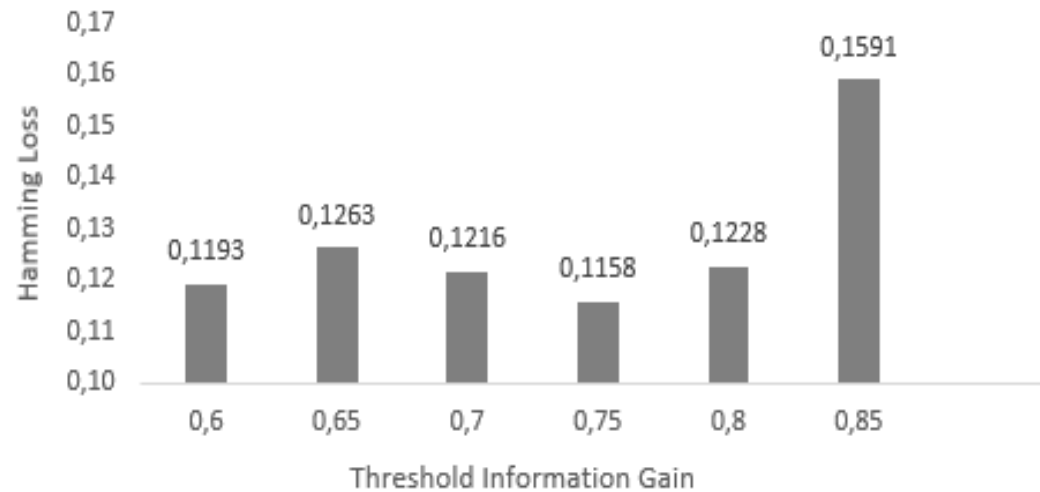




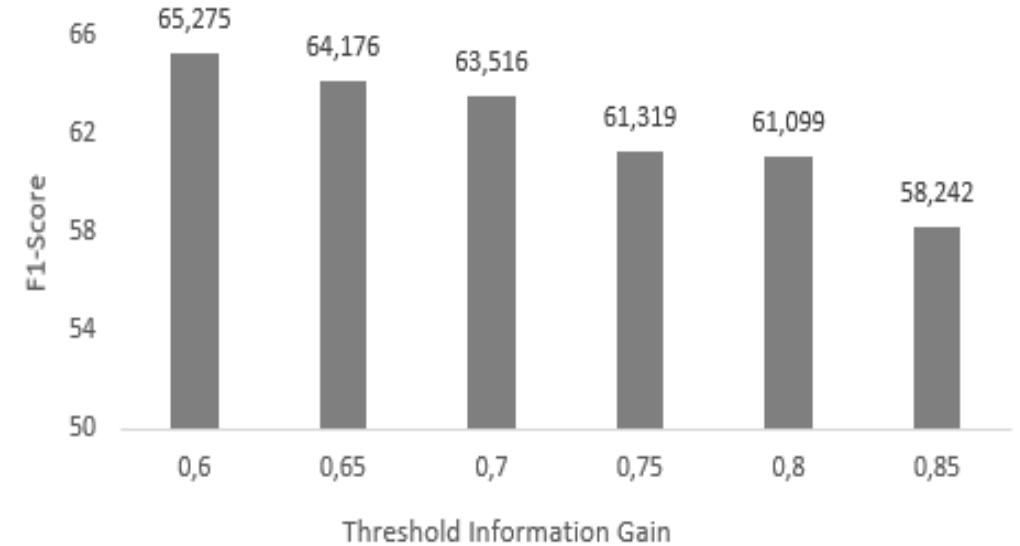
# Evaluasi (cont.)

- Skenario Pengujian dengan Threshold Information Gain

Skenario pengujian nilai *threshold IG* pada terhadap Hamming Loss



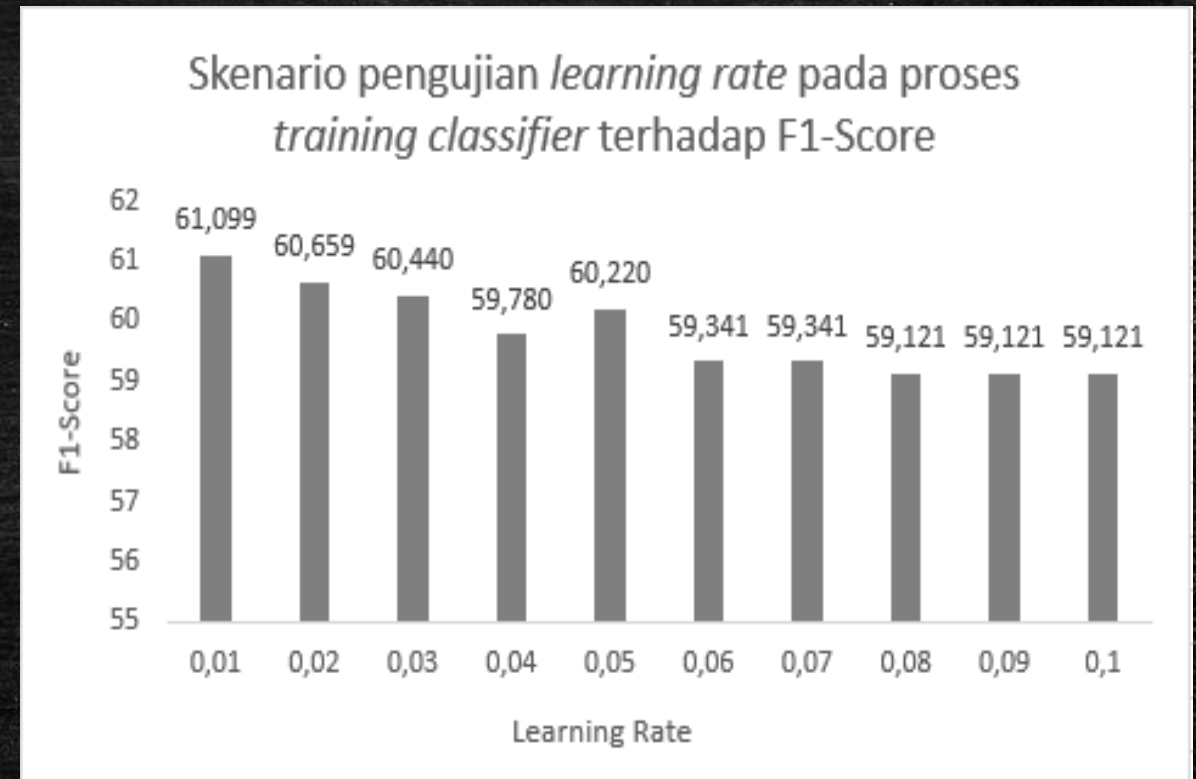
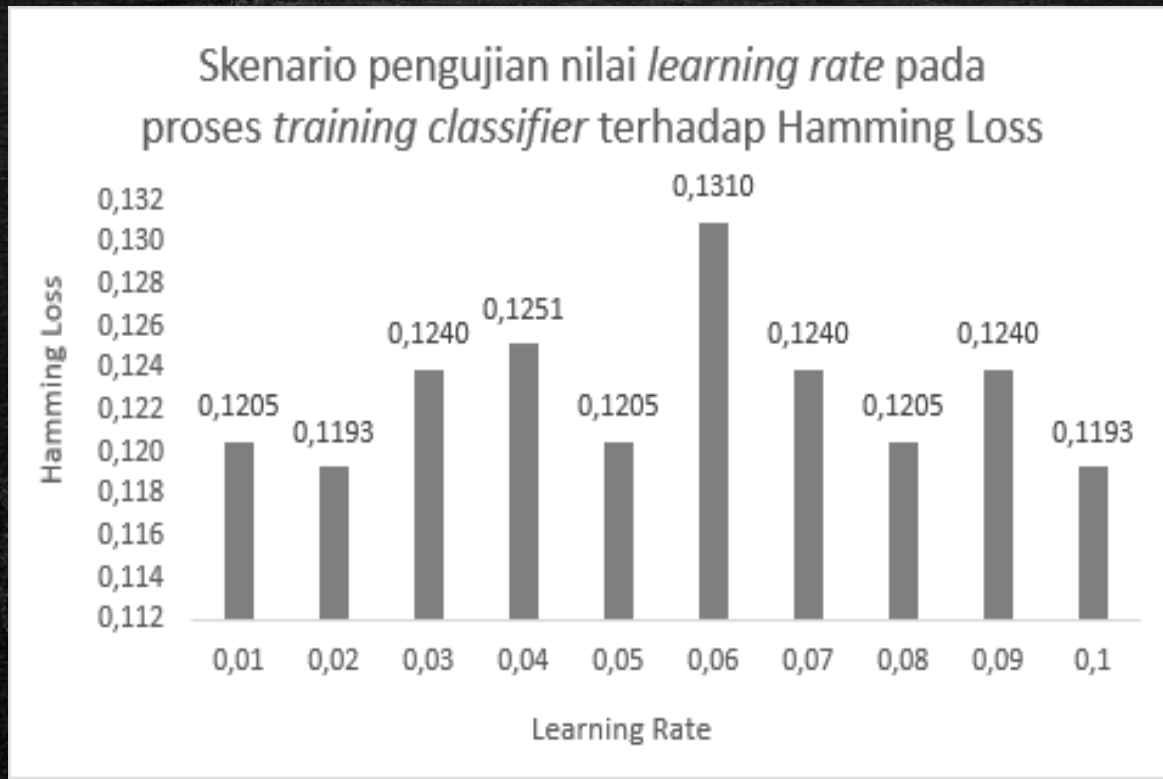
Skenario pengujian threshold Information Gain terhadap hasil F1-Score





# Evaluasi (cont.)

- Skenario Pengujian dengan Nilai Learning Rate





# Kesimpulan dan Saran

---

- Pengaruh penggunaan information gain bekerja efektif pada data multi-label yaitu sebesar 88.42% data multi-label berhasil diklasifikasikan dengan benar dengan menggunakan nilai threshold information gain sebesar 0.75 dibandingkan dengan nilai threshold yang lain.
- Pengaruh proses stemming mendapatkan hasil lebih baik pada data single label namun tidak pada data multi-label. Hal ini karena proses stemming akan menghilangkan ciri khas pada data multi-label.
- Untuk penelitian selanjutnya dapat lebih memperhatikan label data khususnya pada data single label. Selanjutnya adalah mencoba menganalisa makna kata yang ada pada data multi-label



Terima Kasih

---