# 1.2.1

A.

«sansynlighet» er besvart med en presisjon på 3 desimaler.

Pga begrensinger i Excel er det ikke mulig å få alt i en graf.

|  |  |  |  |
| --- | --- | --- | --- |
| Fakultet | 2008 | 2014 | Sansynlighet (basert på 2014 data) |
| Helse -og idrettsfag | 1420 | 1829 | 0,173545877 |
| humaniora og pedagogikk | 1182 | 1525 | 0,144700636 |
| kunstfag | 293 | 420 | 0,039851978 |
| teknologi og realfag | 1337 | 2166 | 0,205522346 |
| lærerutdanningen | 1158 | 1506 | 0,142897808 |
| økonomi og samfunnsvitenskap | 2398 | 3093 | 0,293481355 |
| totalt | 7788 | 10539 | 1 |

B. basert på sannsynlighetsfaktoren og hva vi har lært om komprimering vil du få minst informasjon dersom du får vite på at eleven tilhører fakultetet for «økonomi og samfunnsvitenskap».

C. Huffman treet er designet så den er «instantaneous» med opptil 5 bit kode, og om mindre enn 5 bit informasjon er gitt vil 0 være siste siffer i den binære koden for et fakultet.

1

Teknologi og realfag. 10

Kunstfag. 11111

Lærerutdanningen 11110

1

1

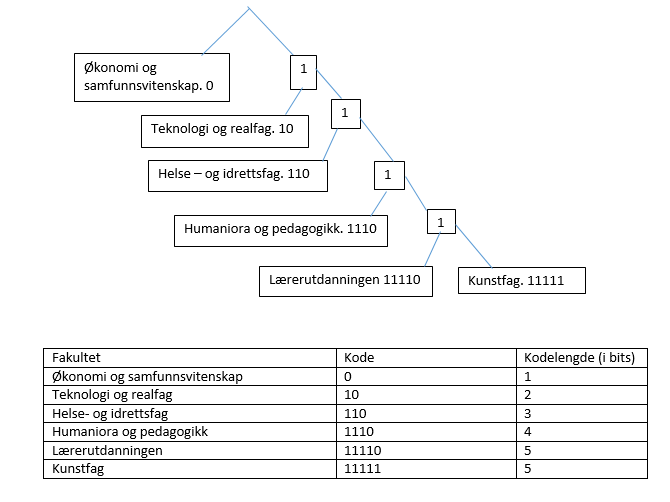
Humaniora og pedagogikk. 1110

Helse – og idrettsfag. 110

1

Økonomi og samfunnsvitenskap. 0

|  |  |  |
| --- | --- | --- |
| Fakultet | Kode | Kodelengde (i bits) |
| Økonomi og samfunnsvitenskap | 0 | 1 |
| Teknologi og realfag | 10 | 2 |
| Helse- og idrettsfag | 110 | 3 |
| Humaniora og pedagogikk | 1110 | 4 |
| Lærerutdanningen | 11110 | 5 |
| Kunstfag | 11111 | 5 |

(akkurat samme som ovenfor, bare som bilde)  


Noen som har noe forslag til how the hell vi kan forbedre dette? For er rimelig sikker på at en kode definert som fast 3 bit lang vil gi bedre komprimering. Noe er veldig feil…

D.   
Gjennomsnittslengden for meldingen med 100 tilfeldig utvalgte studenter blir 271,77 bits. Dette er kontrollregnet utfra mengden bits for hvert fakultet, og dens respektable sannsynlighet, summert sammen.

Utfra Shannons formel for utregning av entropi er entropien 2,41 for dette tilfellet. Altså gjennomsnittlig forventet bit per student er 2,41 bits. I og med at koden er suboptimal (dårlig komprimert) virker dette riktig.

E.  
Algoritmen vil splitte opp det som skrives for hvert femte tegn (3 i den andre versjonen) og ved en while-løkke søke gjennom for å se om de passende kodene finnes i det som ble skrevet inn. Det vil deretter bli printet de passende fakultetene.