Curs: Analiza statistică a datelor

Tema: Probabilitatea condițională și entropia

Sarcina: De calculat entropia medie per simbol pentru textul dat în varianta respectivă

Exemplu.

Textul dat:

***cucurigu, boieri mari, dați punguța cu doi bani!***

Sarcina: de calculat entropia simbolului **u** și entropia lui condiționată cu litera precedentă **c**.

Rezolvarea: In textul dat sunt 48 caractere (cu spații) și 6 litere **u**.

Probabilitatea literei u în textul dat este P(u) = 6/48 = 1/8 = 0,125

Entropia, respectiv H(u) = -P(u) log2(P(u)) = -0,125 log2(1/8) = -0,125 (-3) = 0,375

Astfel, entropia literei u este 0,375

Entropia condiționată este calculată în baza probabilității condiționate.

H(u/c) = -P(u/c) log2(P(u/c))

Probabilitatea condiționată se calculează cu formula

P(u/c) = P(c,u)/P(c)

In textul dat sunt 3 cazuri când litera u este precedată de litera **c**. P(c,u) = 3/48 = 1/8

Probabilitatea literei c în textul dat este P(c) = 3/48 = 1/8 deoarece apare în textul dat doar de 3 ori.

Respectiv, P(u/c) = P(c,u)/P(c) = 3/8 / 3/8 = 1

(Explicația: în textul dat după orice litera c urmează litera u, ce înseamnă că dupa c cu siguranță va urma u; probabilitatea 1)

Entropia, respectiv H(u/c) = -P(u/c) log2(P(u/c)) = -1 log2(1) = 0

În cazul dat probabilitatea este 1, ce înseamnă eveniment sigur și nu este nici o incertitudine. Entropia, ca măsura incertitudinii, în cazul dat este egală cu 0.

Variantele

1. Frumos e muntele ce-nalţă spre ceruri fruntea lui semeaţă,

Litera u, și litera u precedată de litera r.

1. Ridică vocea orice frunză şi-n om se schimbă-orice copac.

Litera i, și litera i precedată de litera r.

1. Tu ce ridici spre ceruri braţe visându-ţi dulcea poezie,

Litera i, și litera i precedată de litera r.

1. Nici verzişori frumoşi ca zmeii, nici poteră şi nici haiduci.

Litera i, și litera i precedată de litera ș.

1. Voinicul dacă pe sub seară şi-adăposteşte-a lui comoară

Litera ă, și litera ă precedată de litera r.

1. Pădure, cadrul tău convine la bucurie şi durere

Litera u, și litera u precedată de litera d.

1. Cu gândurile mele singur aş vrea ca să mă pierd în tine,

Litera r, și litera r precedată de litera u.

1. Ici şi colo, pe afară,  Zboară câte-un fulguşor.  Dar e zi de primăvară: Soare, pune-ţi Mărţişor.

Litera o, și litera o precedată de litera ș.

1. Obosită rău de tot, Furnicuţa s-a oprit Pe un fir de păpădie

Litera ă, și litera ă precedată de litera p.

1. Păpădia-ncetişor Se dă huţa: na, ni, na, furnicuţo, somn uşor.

Litera a, și litera a precedată de litera n.

1. Deodată, furnicuţa Observă că zboară-n sus

Litera a, și litera a precedată de litera ț.

1. Firul cel de păpădie De la locul lui s-a smuls

Litera l, și litera l precedată de litera u.

1. Lumea largă – hopa-hop – Vrea s-o vadă

Litera o, și litera o precedată de litera h.

1. Pe drum trece – dur-dur-dur – O maşină

Litera u, și litera u precedată de litera d.

1. Mititica Vrea acasă – iute, iute – La mămica

Litera u, și litera u precedată de litera i.

1. Şi răţuşca – hop-hop-hop – În ogradă

Litera o, și litera o precedată de litera h.

1. Până creşte, o să stea Doar în curte

Litera e, și litera e precedată de litera t.

1. Un cireş în primăvară Nu e pom, e doar o floare uiaşă, uriaşă, ca zăpada la culoare

Litera i, și litera i precedată de litera r.

1. Plouă, plouă – pic, pic, pic, Să crească spicul voinic.

Litera c, și litera c precedată de litera i.

1. Pic, pic, pic – la geamul meu, cu spicul să cresc şi eu

Litera i, și litera i precedată de litera p.