ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΗΠΕΙΡΟΥ

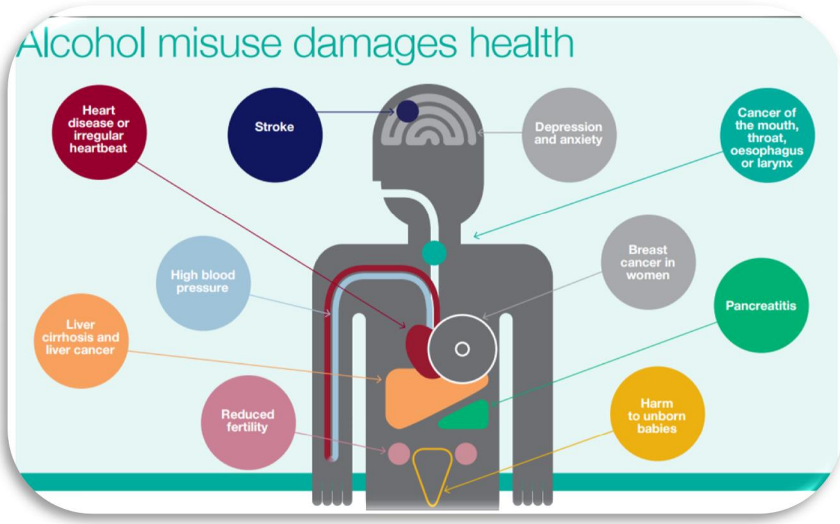
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ Τ.Ε.

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ

ΤΕΙ ΗΠΕΙΡΟΥ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

«ǰ Λ¢ Ɇ¢ ɀ ȀŪɃɋɋ ʔ ɇȷ ʔȽʔɃɀ ¢ Ɇ¢ Ɇ ũȴ ɇ¢ ʔ ɄΛɃȸ ȀɊ¢ ɇ¢ Ɇ Ƭȷ ɇȷ ʔȷ ȿ ɋ Ɇ¢ Ɇȷ ȿ ƬɃɃȿ ȷ ɄɃ ū ɃȾ¢ ɇ ȀɆ»



ƬǗíŸæg ǳæŬůďŬ Α.Μ. 9989

Επιβλέπων καθηγητής

Νικόλαος Γιαννακέας

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ Τ.Ε.

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

«ĵ Ʌ¢ Ɇ¢ ɀ ȺŪɃ ȹɋ ʔ ɇ ȷ ʡ Ƚʔ Ƀɀ ¢ Ɇ¢ Ɇ ũȽȷ ɇ ¢ ʔ ɄɅɃ ȸȿ ȺȽ¢ ɇ ¢ Ɇ Ⱦȷ ɇ ȷ ʔ ȷ ȿ ɋ Ɇ¢ Ɇ ȷ ȿ ȾɃ Ƀȿ ȷ ɄɃ ū ɃȽɇ ¢ ɇ ȺɆ»

Ⱦ Ůí Ÿɘɕg ȸɚŬůɗŬ Α.Μ. 9989

Επιβλέπων καθηγητής

Νικόλαος Γιαννακέας

- Άρτα 2017 -

ÅgɢǓį ϑůǗ dǓį ɛ Ÿ Ǘɛɜ ə. ũϑǓɜɜǓəɏǓ ʔ ϑɧəǓɛ ɔϑǓ Ǘɖɜ Ů ɞϑɞǔɛɛɖǗəɐ ůɡɜǓį ɔǓůɑǓ əǓə
əǓɗɞǔɐɔɖůɖ əǓǗŸ Ǘɖ ǔϑŸį əǓϑǓ Ǔə´ ɧɜɖůɖʃ Ǘɖʃ Ǔį ɔǓůɑǓʃ.

¢ ´Ûğcəİəe Ûİɔ˘ÛûıÛ´ɵg gəɵ´ɵæůÛɛÛİɣcÛəıdÿɛÛ˘Ûłɜ cİɣeůd ɛÛűłĥŭɤɜ ÛÛɜəəĥɛdůdgſɔəÛ˘Ûłɜ ´İİĥbəÛɣd Ûłgſ əÛÛÛɜŸəɤůdgſ Ûəəɵĥə· Ü ĥ ű¢əÛłÛÿgſ. Æəɵ´ĥgſ Ûłgſ ÛİɔÛûıÛİgſ ÛœÛə ɜÛ ŭəÛİÛğɜdſɵei ɜ ÛÛÛŭÛűÛÛɵ´ɵdÛɤſɵei ɜ Ûɛ¢gſ ɜÿɵ¢gſůÛłɜ cİɣeůd Ûɛ¢g Ûəəɵĥə·, əÛəɜÛÛɜÛŭıÛ¢cſÛıd ɛÿəÛÛÛıd Ûɛ¢g ű˘Ûəɵ¢ɛÿɜɵ¢g ůÛɛ¢gſűɵəÛłÛÿgſÛİÿəÛɛbŸſɛəÛİgſÛə´Ûœı˘Ûğůıdgſ. ũəÛÛɛɜ ÛəɤÛÿİɣɤ ůəɵ´ĥ ŭdɛæ¢İɣɔɐcſdəÛÛİɣɤÛłɛÛÛɛəĥɔəæ bÛŭəůɛÿəɵ ůÛŭəÛſɜɐ ÿİÛğɜÛ´ɵg ŭəÛəɐ¢cſdɛÛ˘ĥ´ɵİÛɛɔŸəɵ¢gſ ÛİÛğɜdÛÿgſ, ɛÛ ÿ¢ű˘Ûıd ůÛłɜ cİɣeůd Ûəəɵĥə· Ü ĥ ɛÛcſdÛÿgſ ůÛłɜ Ÿ¢İÿ˘ÛɛɔÛəcÛ. ÆÛÛ ÛİɤÛłɛÛÛɛəĥɔəÛ´ɵg ůgəəəÿ¢cſdəÛɜ ÛſıÛİɛĥůÛłəÛɜ ɛÿcſɵŭɵ¢ÛÛəæĥɛdůdgſɛÛůəɵ´ĥ ɜÛÛəÛ¢cſÛıd Ûɛ´ɵůɵůÛĥ Û əÛğɵ¢cſÛgſ Ûłgſ ´İĥbəÛɣdgſ Ûłgſ cİɣeůdgſ Ûəəɵĥə· ɛÛ bŸůd Ûɛgſ Ü ÛɜÛɛůÛəgſ ÛɤɜÛİɤÛłcſdÿɜÛɤɜ. ÆÛɛ´əÛůıɵæ ÛɛgſÛİɔÛûıÛgſůgɔəÛɜÛİı cſdəÛɜ 133 ÛİɤÛłɛÛÛɛəĥɔəÛÛəı ůıɛűɤɜÛ ɛÛÛÛ Ü ɵÛÛəÿůɛÛÛÛ Ûłgſ, Û əÛÛıɵ¢cſdəÛ´ɵůɵůĥ ÿɤɾgſ əÛɔ 80% ůÛłɜ ´İĥbəÛɣd Ûłgſ cİɣeůdgſ Ûəəɵĥə·

ÆÛɛ ´İı Ûɛ əÛűŸəÛəæ Ûłgſ ÛİɔÛûıÛgſ ɛÛgſ cſÜ ÛůɵɵəæcſdɵeıɛÛÛɔÛəæŸɛÛÛɛɜ Ûəəɵɵəæůɛĥ ůÛɛɜ ´əæcſdgůɛĥ, cſÜ ůgɔəÛɜÛİı ůɵɵgɛÛ ůÛÛÛɛůÛəæŸ ůÛɛæɵÛůÜ əÛɔ cſÜ ÛůıÛ¢ŸůɵɵgɛÛ ůÛɛɜ Ûəəɵɵəæůɛĥ ůÛɛɵgſ ɜÿɵ¢gſ. ÆÛɛ ŭÛıÛÛİɛ əÛűŸəÛəæ cſÜ ÛəÛÛİÛİcſdɵeıɛÛ ůÛÛ ÛğůgÛůgſ ɛÛűłĥŭɵ¢gſ ɛcſdÛÛəæɵgſ ɛŸ¢cſdůcſdgſ ĥ´ɤgſ ɔəÜ´ÛİɤŸŭÛəæɛÜ ÛɛɜTˆəɵɜÛæəĥÛİɛ ɔÛűÛɛɜÜ əÛɔ ÛÛŭ üÿɜÛİɤÜ Ü ɵűŸůÛɤɜ. ÆÛɛ Ûİcſdɛ əÛűŸəÛəæ Ûłgſ ÛİɔÛûıÛgſ ɛÜgſ cſÜ ɔɜɤɾİůɵ¢gɛÛ Ûɛ ´ÛİɕbŸəəɵɜ Weka Ûɛ ɵ´¢ɵæ cſÜ ɾİcſdůɛæ´ɵæůɵ¢gɛÛ ůÛÛ Ü ɵÛÛəÿůɛÛÛÛ Ûɛ¢g ÛİɤÛłɛÛÛɛəɔɔæg ɛÜgſ. ¢ɛ ÛÿÛÛİÛɛ əÛűŸəÛəæ ÛűɵəİŸÛłɜ gəæ´ɵcſdůcſ ´ÛĥİÜɛŸÛɤɜ. TˆÛɔ ÛÿəæɵgſůÛɛ ´ɤɛ´Ûɛ əÛűŸəÛəæ cſÜ ÛůɵɵəæcſdɵeıɛÛɛÛÛÛ ůÛÛÛÛıÛÛəæŸůÛɛæ¢ÛÜŭÛæɛÛÛɛgſəÛcſdıgſəÛɔɛÛÛÛÜ ɵÛÛəÿůɛÛÛÛ ÛÛÛɜəəĥɛdůdgſ.

.

The current work aims the investigation of the multifactorial problem of alcoholism during academic studies. Machine learning techniques have been employed to compute the possibility of alcoholism in students. Our work is based on a previous work from a Portuguese research group, where several factors have been investigated for alcohol use in school. To gather data from students, an on line questionnaire is employed and shared in social media. Four classification algorithms have been used to classify the students regarding their use of alcohol. A prediction accuracy about 80% is obtained in a dataset of 133 students.

In the first chapter of our work will generally deal with alcoholism in the population, we will gather statistics and will focus on alcoholism among young people. The second chapter refers to intelligent machine learning methods such as the K-nearest neighbor and decision trees. The Weka environment, which is used for mining the data of our questionnaire, is presented in the third chapter, while the fourth chapter focuses on the implementation of experiments. Finally, statistical sample data and the classification results, are presented in the results chapter.

ɀ ȺʔqʔȠ

ⅎ Au ȷ ȿ ȷ ⱠƁ Ɪ

ȼ Ɓ ɄɅƁȸȿ ȼ ȥ ȷ ȼ ƁɈ ȷ ȿ ȾƁƁȿ ⱠȺȥ ƁɈ

1.1 ȷ ɘɘ ɞɞɘɘ ůɛ ɧɠ

ȥ ŮŰɞɜ ɧȝɞ Ŭɘɘɞɞɘɘůɛɧ ŬɜŬŰŬ ɧɛŬŰŰŮŮŮɛɘŬ´ Ŭɖɞɘɞɔɘɐ ůɕɏůɖ Űɖɜ ɞ´ ɞɗŬ üɨ ɜŬŰŰɘ ɜŬ ŬɜŬ Űɨ ɜŮɘ ɏɜŬ ƳŰɛɛɞ ɛŮŰɛ Ŭɘɘɞɧɘ. Ɓgůɘ ŬůŰɘ Ƴ ɛɘƳɛŮɔɘŬ Űɖɜ ´ɘɐȝɖ ŮɘƳȝ Űɖůɖ Űɛg ŬŰɧɛɞg Ŭ ɧ Űɛ Ŭɘɘɞɧɘ. ƁɘŮ ɘ Ű ůŮɠɔɘŬ Űɖɜ Ŭɜɖȝ ı ´ɜɖ gɔŮɖŬ ŮɞŬɘ´ ƳȝŬ´ ɞɘɘɏɠ ɘŬɘ üɘŬɘȝɞɞɜŰŰɘ ůŮɣgɞɘɏɠ ɘŬɘ ůɤ ɛŬŰɘɏɠ. Ɓ Ŭɘɘɞɞɘɘůɛɧɠ ɛ´ ɞȝŮɑ ɜŬ ´ Ŭȝ ɞgůɘƳůŬɘ ŰŰ´ ŬȝŬɘƳŰɤ ůgɛ´Ű ɛŬŰŰ:

- ***ɑɘŬɘŬɐ ´ɧɖɞ***
- ***ɉ Ƴůɘɛɞ Ůɘɏɔɞɞg***
- ***ū gůɘɐɐ ŮɘƳȝ Űɖůɖ***
- ***Ⱥɖɘůɛɧ***

Ⱥɘɘɧɜ Ŭ** 1.1: **ȷ ɘɘɞɧɘ

ȳ ůɞɜ Űɞɞȝ Ƴ Űɖɜ üɘŬɘŬɐ ´ɧɖɞ ŬɜŬŰŬ ɧɛŬŰŰŮŮŰɖɜ ůŰɞŬȝɐ üɑɣŬ ɘŬɖ ɠ ɘŬɘ Űɖɜ ɘŬɢŰƳȝŬ ɔɘŬ ´ɞŰɧ. ȷ ´ɧ Űɖɜ Ƴɘɘɖ ´ɘŮȝȝƳ Űɛ ɢƳůɘɛɞ Ůɘɏɔɢɞg ɏɢŮɘ ɜŬ ɘƳɜŮɘ ɛŮŰɛ ɢȝɞɜɘɧ üɘƳůŰɖɛŬ ɑɖɘŬüɐ Ŭ ɧ Űɖɜ ůŰɞɛɐ ´ɞg ɘƳ ɞɘɞɠ ɜŮɘɜƳŮɘ ɜŬ ´ɞŮɘ ŮɞŬɘ

Ŭ Ǘ ůŰŰεŬŰǗůǗ ¢ ǘgůǝɐ ǓǗŰůd Ŭ ɞŰǗǗ ŰŰεŬŰŸ3Ů3Ŭ´ɶɞɡ3. ɀ Ǘ ǝŸŬ ɦ ŰŬůɡε´Ű εŬŰŬ ǗǗŬǗŰŬǗɐɠ:

- ***?ŬǗŰǗŬ***
- ***ǺǗǎǗ ɤ ůd***
- ***ȼɉɏεɞɡǝɞ***
- ***ȯ ɔɞɞɠ***

ȼɏǝɞɠ ɞ ǗǗůεɦɠ ŭdǝǗɐ d ǓŸǝd Ǘɡ Űɦεɞɡ 3Ŭ ǝŰŰǓŬǗ ůǗ ´ɞǝɨ εŮǗŸǝǗɠ ´ɞůɦǗǗŰǗŬǝɞɦǝ.

1.2 Ƀ Ŭǝɞɞǝǝůε ɦɠ ůǗ3 ´ǝdɗɡůε ɦ

ȼŬ ŰǗǗŰŰŬ ɢɉɦ3ǝŬ ɏɢǗ ´ǗŬǗɉdǗ ɏ3Ǘ3Ŭ d ǝŰŰǓŸǝɤ ůd ǓǝɞɦǝůŮ ´ɞǝɨ εŮǗŸǝŮɠ ´ɞůɦǗǗŰǗ. ɆŰŰǗǗůǗǝɏɠ εŮǗɏŰǗ ɏŭǗǗŬ ɦǗ 3,3 ǗŬŰǗεεɨɉǝŬ ǓdɉɩΊ´ɤ3 ´ǗǗŬɞɞɡ3 Ŭ ɦ Ǘ Ŭǝɞɦǝ; ɞɡůǝŬůǗŸ 1 ůǗɡɠ 20 dǓŸǗɡɠ ɞŭǗǝɞ3ŰŰ ůǗ ´ɞǗɦ. Ɇɨ εǘɤ 3Ŭ εŮɏdŮd Ǘɡ ɄǓǝɦůεɞɡ ɃɉɔǓǝůεɞɨ Ɉ ɔŮǗǗ (ɓǝ. Ʉ.Ƀ.Ɉ) ´ǗǝůɦŰǗɞǝ Ŭ ɦ 200 ŰůdŮǗɠ ůɡ3ŭɏɞ3ŰŰ εŮǗ3 ǝŰŸɢɉdůd ɞɞɞ´3Ǘ εŰǗ ǗŭΊ 3 ´ɞǗ 3. ȼɞ 2012 ´ǗɉǗǗɉɐddǝŬ ɦǗ d ɡ´ǗǝŰŰǓŸǝɤ ůd Ǘɡ Ŭǝɞɦǝ ´ɉɞŸǝǗŮ ´Ǘά ɞɡ 3,3 ǗŬŰǗεεɨɉǝŬ dǓŸǗɡɠ ůǗ3 ǝɦůεɞ ǗǝŰŸεǗŬǗŰǗd ůŮůɢɏůd εŬǗ 2005 ´ɞɡ ɞǝ dŸ3ŰǗǝ ŭǗ 3Ů Ǘ3ɞɨ3 Ű 2,5 ǗŬŰǗεεɨɉǝŬ. ¢ ǝŰŰǓŸǝɤ ůd Ǔǝɞɦǝ ŭǗ ɡǘǗŰŰǗ εɦ3ɞ ůŬ´ǝɞɨ ůǝǗɠɢɩ ɉǗɠŬǝǝŸǝǓůŮɢɩ ɉǗɠ´ɞɡ ǓǗε Ű ´ɶɞ3Ű ůɞɓǗŸɞǝɞ3ɞεǝŸ ´ɉɞɓǝɐεŰŰ.

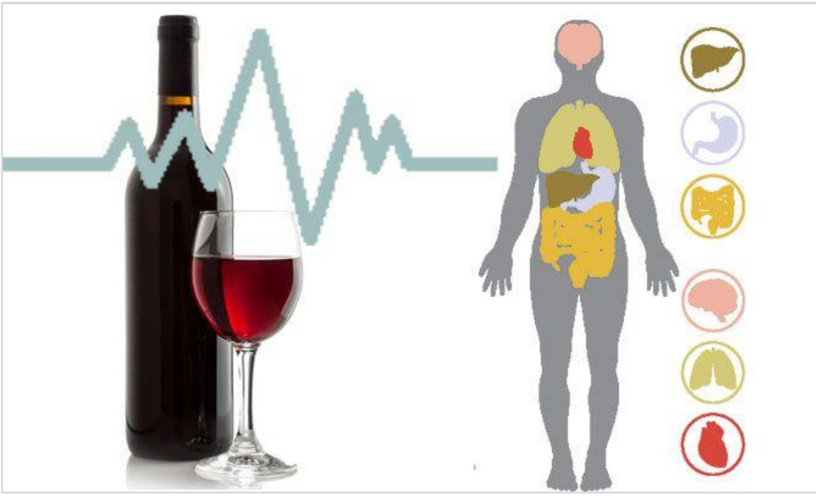
ǺǝɦɜŬ1.2 Ƀ Ŭǝɞɞǝǝůε ɦɠ

ɞɨ

Ɇɨ ɛuɤ ɜƯ ɛƯuƯƯɑuƯɘɣɡ ɛƯɘɣƯɑɡ ɖ Ⱥɡɟ ɩ ˊɖ Ư ɞƯƯɘƯɑ Ưɖɜ ɛƯɘƯɘɨ ƯƯɟ ɖ ɐˊƯɟ ɞ ůƯɖɜ ɚƯƯƯɜƳɚɤ ůɖ Ưɘɚɞɧɚ. Ʉɕ ůɡɔɚƯɘɟ ɛɣɜƯ:

- ✓ ȾƯƯƯɜƳɚɤ ůɖ Ưɘɚɞɧɚ ƯɜƳ ƳƯɘɛɞ ůƯɖɜ Ⱥɡɟ ɩ ˊɖ 9,3 ɚɖɟ Ư
- ✓ ȾƯƯƯɜƳɚɤ ůɖ Ưɘɚɞɧɚ ƯɜƳ ƳƯɘɛɞ ůƯɡ ¢ Ʉɟ 6,6 ɚɖɟ Ư
- ✓ ȾƯƯƯɜƳɚɤ ůɖ Ưɘɚɞɧɚ ƯɜƳ ƳƯɘɛɞ ůƯɘɜ ȾƯɜƯɑƳ 6,1 ɚɖɟ Ư

Ƀɡůɘ ƯůƯɘƳ Ư ɧ 320 ƯɘƯƯɘɛɛɨ ɟ ɘƯ Ưɘɡ ůɡɜɞɚɘɞɨ ˊɚɖɡůɛɞɨ ɣɜƯ ƯɘƯɟ ƯƯɘƳ ɛƯɘƳɚɞ ɛɣɟ ɞɡ Ưɘɡ ɡˊɞůɣɟ Ưɘ Ư ɧ Ưɡ Ư ɘ Ư ůƯɡ Ưɘɡ Ưɘɚɞɞɚɘůɛɞɨ ɚƯɘ ůɡɔɚƯɘɟ ɛɣɜƯ 84 ƯɘƯƯɘɛɛɨ ɟ ɘƯ ƯɘƯɘ ƳɛůƯ ƯɘƯɟ Ưɖɛɣɜɞɘ Ư ɧ Ưɘ Ưɘɚɞɧɚ, Ưɘɩ ƯƯ 80 ƯɘƯƯɘɛɛɨ ɟ ɘƯ ɓɑ ɜɞɡɜ ɚƳ ɞɕ ˊɟ ɧɓɚɖɛ Ư ˊɞɡ ůɡɜůɣƯƯƯɘɛƯƯɘɜ Ưɘɚɞɞɚɘůɛɧ.



Ⱥɘɧɜ Ư 1.3 : ɟ ɚɚɞɧɚ ɚƯɘ ɡɔƯƯ

ɣ ůɞɜ Ưůɞɟ Ƴ Ưɖɜ ɚƯƯƯɜƳɚɤ ůɖ Ưɘɚɞɧɚ ůƯɘɜ ƯɘɚƯůɘɧ ɞɩ ɟ ɞ Ưɖɜ ɞɟ ɞɜɞɚɞɔɘɐ ˊƯɟ ɞɘůɞ Ư ɧ Ưɘ 1975 ɣɤ ɡ ɚƯɘ Ưɘ 1980 ˊƯɟ ƯƯɖɟ ɐɖɖɚƯɣɜƯɘɜ Ư Ưɘ ůƯɘɧɛƯɘɞ Ưɖɡ ɚƯƯƯɜƳɚɤ ůɖɡ Ưɘɚɞɧɚ ɚƯɘ ůɡɔɚƯɘɟ ɛɣɜƯ ůƯɘ 5,3 ůƯƯ 10,2 ɚɖɟ Ư ƯɜƳ ƳƯɘɛɞ. ɆƯɖ ůɡɜɣɞƯɘƯ ɖ ɚƯƯƯɜƳɚɤ ůɖ ɣ ƯůƯůƯƯ 8,7 ɚɖɟ Ư. ɋɚƯ ɘůƯ ɖɖɚƯɧƯɘɖ ɚƯƯƯɜƳɚɤ ůɖ ɛ ˊɨ ɟ Ưɡ Ưɡɜɐɖɖɚ Ư ɚƯƯɘɧɟ ɡůƯ ɚƯɘ ɖ ƯɛƯɘɖƯɣƯ ˊɞůɧƯɖƯƯ ˊɞɡ ɚƯƯƯɜƯɘɩ ɜɞɜƯƯɜ ˊƯɟ ɑ ɞɡ Ưɖɜ ɞɟ ɞɜɘɐ ˊƯɟ ɞůɞ 1990 ɣůƯƯůƯ ƯƯ ƯƯɘƯɟ ƯƯɖƯ ɞɟ ɧɜɘƯ ůƯƯ 40 ɚɖɟ Ư ƯɜƳ ƳƯɘɛɞ. Ⱥˊ ɑůɖɡ ˊƯɟ ƯƯɖɟ ɐɖɖɚƯɣɜƯɘɜ Ư Ưɘ ůƯɘɧɛƯɘɞ ɛƯɘƳɚɤ ɜ ˊɞůɞƯɘ Ư ɜ ˊɡɚɜɩ ɜ ˊɞƯ ɜ ɚɧɔɞɡ ɞƳɟ ɖ ɞɡɑůɚɕ ɓɧƯɘƯ ˊɞɚɨ ˊƯɟ ɘůůɧƯƯɟ ɞ Ư ɧ ɧƯɘ ůƯɖɜ Ⱥɡɟ ɤ ˊƯɘɐ ɣɜɤ ůɖ. ɟ ɜɞƯɘ ɜƯ ƯɜƯƯɣɟ ɞɡɛƯ ɧƯɘ ɖ ȺɚɚƳƯƯ ɣɟ ɞƯƯƯɘ ˊɟ ɩ Ưɖ ůƯ ɚƯƯƯɜƳɚɤ ůɖ Ưɘɚɞɧɚ ůƯɘɡɡ ɜɣɞɡɡ ɖɚɕɖƯɡ 15 ƯƯ ɜ ɚƳƯɘ ˊɞɡ ɖƯ ƯɜƯɘɨ ůɞɡɛƯƯɘƯƯɜɣůƯƯɟ Ư ůƯƯ ɧɛƯɘɞ ɡ ˊɞɚƯƳɚƯɘ Ưɖɡ ˊƯɟ ɞɨ ůƯɡ ƯɟɔƯůɖƯɡ. Ƀɚɞɚɚɖɟ ɩ ɜɞɜƯƯɡ Ưɖɜ ƯɜƯƯɞɟ Ƴ ɛƯɡ ůƯɖɜ ɚƯƯƯɜƳɚɤ ůɖ Ưɘɚɞɧɚ

ɖɛƳɜɕɡɛŮɧǿɖ ɢɩ ɟŬɛŬǵ ŮɑŬɜ´ɟɩ ǽɖ ůŮɜɕɖǾɖŬ ɖ. ¢ ´ǚɟŬ Ƴɜɤ ɕɡůɖŬɛ´ɕɟŮɑɜŬ´ɟɕəŬəɏůǿ´ɕəɨ ůɖɛŬɜǿəƳ´ɟɕɓəɐɛŬǾǾ ůŮɕ´ǿəƳɜŮɟŬəŬə´ɞ ůɡɔəŮɘɟɛɏɜŬɜŬɕŭɖɔɐůǿəŬəůŮ Ǿ ǘəɤ ůɖ.

ɟ əɕəɕɡɖŮɑŬɜŬǚɕɟƳɔəŬǽɖɜ əŬǾǾɜƳəɤ ůɖ Ŭəəɕɧə ůŮŮ əŮɛɏɜǵɢɩ ɟǵůɨ ɛǘɤɜŬɛŮ ůǾǾǾůǿəƳůǿɜɢŮɖŬ:

ɟ ɡůǿǚɖŬ

Ǽǽɖɜ ɟ ɡůǿǚɖŬ ´ǚɟŬǽɖɟɐɖɖəŮɧǿ ǾŬ ǾŮəǵɡǾǾɖŬ ´ǚɑɕɡ 30 ɕɟɧɜəŬɖ əŬǾǾɜƳəɤ ůɖ Ŭəəɕɧə əɡɛŬɑǾǾǾəŬ ɧ 11,2 ɏɤ ǵ 9,8 əǽǚ Ŭ əŬɖŬɟɕɨ Ŭəəɕɧə ŬɜƳ Ƴǿɛɛɕ. Ǻ´ɖůɖǵ ůǽɖɜ ɟ ɡůǿǚɖŬɡ´ƳɟɢŮɛŮƳəɖ ´ɕůɧǽɖǾǾ ò Ŭəəɕ´ɧ´ɡǒ´ɕɡ ŭŮɜ ŮɑŬəƳəəɕ Ŭ ɧ ŬɜŬɣɡəǿəɧ ɛŮɢŬɛɖəɧ ´ɕůɕůǾɧ Ŭəəɕɧə. Ǽɨɛǘɤ ɜŬ ɛŮ ůǾǾǾůǿəƳ ůǿɜɢŮɖŬ ǽɖǵ ɄŬəɧůɛəŬǵ Ƀɟ ɔƳɜɤ ůɖǵ Ɉ ɔŮɖŬǵ (ɓə. WHO) ɛ´ɕɟŮɑɜŬ ŭəŬ əŮǿɤ ɖŮɑɧǿ ǿɜ 28,8% əŬǾǾɜƳəɤ ůɖǵ ɛŮŬɨǾǚɖǵ ´ɕůɧǽɖǾǾǵ Ŭ ɧ 60 ɔɟŬɛɛƳɟəŬ Ŭəəɕɧə ůŮɖɛǚɐůəŬ ɓƳůɖ ŮɑŬə Ƴɜŭɟǵ Ůɜɩ ǿɜ 4,35 ŮɑŬəɔɡɜŬəǵ.

ŭŬəəɖŬ

Ǽǽɖɜ ŭŮŬǾǾɖŬǿɡ 1960 ɖ ŭŬəəɖŬ ŮɢŮəŬǾƳəŮɨŬəɐ əŬǾǾɜƳəɤ ůɖ ´ǚɑɕɡ 17,5 əǽǚ Ŭ. ɇəǵ Ů ɧɛŮɜǵ ŭŮŬǾǾɖǵ ɖ ɛŮɤ ůɖ Ŭɡǿɐ ɐǽǿɜ ŮəŬǚǾǿəƳ ůɖɛŬɜǿəɐ ɧ´ɕɡ ɏǘǾǾůŮǾǾ ǾŮəǵɡǾǾɖŬ ɕɟɧɜəŬ ůǾǾ 11,1 əǽǚ Ŭ ŬɜƳ Ƴǿɛɛɕ. Ǽɨɛǘɤ ɜŬ ɛŮ ůǾǾǾůǿəƳ ůǿɜɢŮɖŬ ´ǚɑɕɡ 2 ŮəŬǾǿɛɛɨɟəŬ Ƴɜɖɟɤ´ɕə ŮɑŬə ŮəŬɟǽɖɛɏɜɕə Ŭ ɧ ǿɜ Ŭəəɕɧə Ůɜɩ 5 ŮəŬǾǿɛɛɨɟəŬ ŬɜǿɛǾǿɤ´ɕɕɡɜ əƳ ɕɞ ´ɟɧɓəɖɛŬ ´ɕɡ ůɢǾǾɛǾǾŭə ɛŮ ǿɜ Ŭəəɕɧə (ɓə. əǚǚ əƳ, ɣɡɢəǚǚ əƳ əŬəƳəəŬ´ɟɕɓəɐɛŬǾǾ).

ŭǚɛŬɜɖŬ

ǼǾǾǾůǿəƳ ůǿɜɢŮɖŬ ŭŮɢɜɕɡɜ ɧǿɖ ŭǚɛŬɜɖŬ ɏɢŮǽɖɜ ŭŮ ǚǚɖ ɖɏůɖ ůǽɖɜ ǚɡɟɤ´Ŭɐɐ ɏɜɤ ůɖ ůǽɖɜ əŬǾǾɜƳəɤ ůɖ ɛ´ɨɟŬǵəŬə´ɞ ůɡɔəŮɘɟɛɏɜŬ´ɕůɕůǾɧ ´ɕɡ ǘǾƳɜŮǾǾ 134,5 əǽǚ Ŭ ŬɜƳ Ƴǿɛɛɕ. Ǻ´ɘ əɏɕɜ ɕə ŮəŬǚǽɖɛɏɜɕə Ŭ ɧ ǿɜ Ŭəəɕɧə ŭŮɜ ɜŮ ǚɜɕɨɜ ǾǾ 2,5 ŮəŬǾǿɛɛɨɟəŬəŬəŬǿɛɏǿɤ´ɕəɛŮǾǾ´ɟɕɓəɐɛŬǾǾɕɟɐůɖǵǿɡ Ŭəəɕɧə ɏɟɢǾǾəůǾǾ 6,5 əǽǚ Ŭ. ɄǚɑɕɡЗ 40.000 ɖƳɜŬǿə ǿɜ ɕɟɧɜɕ ɕǘŮəɕɜǾǾə ůǽɖɜ ɡ´ǚɓɕəɐ əŬǾǾɜƳəɤ ůɖ Ŭəəɕɧə.

[illegible] 142,5 [illegible] 9,1 [illegible]

[illegible]

[illegible] 4,9 [illegible] 300 [illegible] 10% [illegible] 3-6%.

[illegible]

[illegible] 5.000.000 [illegible] Dr. V. Egorov [illegible]: «[illegible]».

1.2.2 [illegible]

[illegible]

igůŰgGı gƒɛɦɜɛ Ű̃ɭɜ ŰŰɔŮůgŰŰɑŮ ŭŰɔŰŰŰŰŮ ď ´ɛəɔŰŰɑŮ
ŭůgɜɛ əŮɔ Ű̃ɡgƒ ŭɔŮůŰŸůŮɡgƒ Ű̃ɛg Ůɔəɛɛəɔůɛɛi. ũi+ɟɤ
ůŰŮ 12 ɣŰ̃ď ŰŮ ŮɔɦɟɤɔŮ ŭɛəɔɛ Ÿɛɛgɜ Ű̃ɛ ´ɟɹı Ű̃ɛ ´ɛŰ̃ɟɹɔɛ´i+ɟŮgƒɛ əɟŮůɔɛi Ůɜı ɣɜŮɛɟɦɜɛ
ŮɟɔɦŰŰɟŮ ŭɛəɔɛ Ÿɛɛgɜ ŰŮ´ɔɛ ɓŮɟɔŸ´ɛŰŸ. Á3 ŮɜŰ̃ɑƒůůɔŰŮ ŰŮ əɛɟɑŰůɔŮ ůŰŮ 13 ɜŮůɔɛɛi ɜ
ɜŮ ŭɛəɔɛ Ÿɛɛgɜ Ű̃ɛ ´ɟɹı Ű̃ɛ ´ɛŰ̃ɟɹɔɛ´i+ɟŮgƒɛ əɟŮůɔɛi Ůɜı ůŰŮ 14,5 ŭɛəɔɛ Ÿɛɛgɜ ŰŮ´ɔɛ
ɓŮɟɔŸ´ɛŰŸ. Á´ɔ´ɟɛůɑƒɣŰɤgɑď´ɟɹı Ű̃ď ɛɣɑďɔəŮŰŮŮɔɦɟɤɔŮůɔŮůŰŮ 14,5 ɣŰ̃ď Ůɜı ɔəŮŰŮ
əɛɟɑŰůɔŮ ůŰŮ 15 ɣŰ̃ď. ¢ əŮŰŮɜŸɔɤůď Ůɔəɛɦɔ•ůŮɛɔɟɤgƒ ďɔɔɔɑƒg´ŮɑŮɔ əŮɑɛɟɔůŰɔɦ
ɟɦɔɛ ůŰ̃ɭ ɛŮŰ´ŮɔŰŮ ůgɛ´Ůɟɔűɛɟ Ÿɛɛg ŮŰɦɛɛg əŮɑɦŮɔ ŮɑƒŮŸɛŮɔ ŮɔəŰíŮɔɔɣgƒ
´Ůɟɔɛɛgƒ ɦ´ɤgƒ ɛ ´ɟɹɛɛŰŰɤ´ɔŮɛgƒ ű̃ɔɛɦƒg ´ɛg əŮɑɛɟɑɛŮɔ Ű̃ɭɜ ɔɛɣɑď Ŭ ɛűŸůŰɤɜ.
Á´ɔ´ɔɣɛɜ ŮɑɛŮɔɔŭɔŮŰŰɟɛ Ű̃ɛɜɔɔɦɔɔŮŰ̃ɛ ŮɔŮ ŮgůůɦɟɛŮɔɛ ɜŮgɟɔɔɦůi+ůŰ̃ɭɛŮ.

ÆŮ ´ɟɦůűŮŰ̃ď ɣɟŮgɜŮ Ű̃ɛg Áɟ Ůgɜɑď Ű̃ɔɔɛi ŬŮɜŮɔůŰ̃ɭɛŮɔɛi Ł3ůŰ̃ɭŰ̃ɛi+Ű̃ɛg CŮgɛɜɔɛƒg
Ɉ ɔɔŮɑɔɛƒg (ɓɔ•. ÁŬ+ŁCŮɟ) ´ŮɟŮŰ̃ď ɟɛďɑďəŮɣɜŰ̃ɛɜɛ Ű̃ɛ űŮɔɑɦɟɛŮɔɛ əŮŰŰɜŸɔ•ůɑƒg Ůɔəɛɦɔ•
Ůűɛɓɤɜ ɛŮɑďɑŰí ɜ. ɟ ɜŮɔgŰɔɑɦŰŰɟŮŮŰŮ ɛŰŰɔ•ɣůɛŮŰŰŮŰɤɜ Ůɟůgɜı ɜ ɣüŮɔɑŮɔ ɦŮɔ1 ůŰ̃ɛggƒ
7 ɛŮɑďŰɣƒ´ɶŮɔ´Ůɟɑ ɛg 2 ɛŮ3 űɛɟɣgƒ Ű̃ɭɜ ŮɔüɛɛŸüŮ. ¢ɛ 42% üďəŮűɛ ´Ůɟɑ ɛg ɛɔ
ɛɔůɛɑ ɛŮɑďŰɣƒ əŮŰŰűŮůɔɛgɜ ůŮ g´Ůɟɓɛɔ•ɔɛ əŮŰŰɜŸɔ•ůď Ůɔəɛɦɔ•. ¢ɛ 15%
əŮŰŰɜŮɔı ɜŮɔ Ůɔəɛɦɔ´Ÿɜɤ Ű̃ɦ 3 űɛɟɣgƒ Ű̃ɛ ɛɒɜŮɔı Ű̃ɛ 14,9% Ű̃ɤɜ Ůɔɛɟɔɹ ɜ əŮɔ Ű̃ɛ
9,5% Ű̃ɤɜ əɛɟɔɑŰůɔɹ ɜ ɣɛɛgɜ ɛŮůi+ůŮɔ ɛɣůŮůɔ ɛ Ű̃ɛ ŰŮɔŮůgŰŰɔɛ ɛɒɜŮ. ɟ ´ɦ Ű̃ɛ 1999 ɣɤɡƒ Ű̃ɛ
2007 ´ŮɟŮŰ̃ɟɹɛďɑďəŮŰi+ɜďůď Ű̃ɛg ´ɛůɛůŰ̃ɛi+ Ű̃ɛg Ůɔəɛɦɔ•əŮŰŰɜŸɔ•ůďɡƒŮ ɦ Ű̃ɛggƒɜɣɛggƒ
´Ůɟɑ ɛg 10 ´ɛůɛůŮŮŰůƒ ɛɛɜŸüŮɡƒ. Ł3ɛ ůgɔəŮɔɟɛɛɣɜŮ Ű̃ɛ 1999 Ű̃ɛ ´ɛůɛůŰɦ ɣúŰŰɜŮ
´Ůɟɑ ɛg Ű̃ɛ 31% Ůɜı Ű̃ɛ 2007 ´Ůɟɑ ɛg Ű̃ɛ 41%.

1.3 Ł3ɟɦɓɔ•Ůɤď Ű̃ɛg Ůɔəɛɛəɔůɛɛi ůŰ̃ɛggƒɜɣɛggƒ

¢ɛ ´ɛŰɦ ɛ´ɛɟŮɑɜŮ ɣGŮɔ əŮɔɦ ŮɜŰɑɔŰɡ´ɛ ůŰ̃ď ɛɤɛ ɛŮɡ. ÐgůɔŮůŰɑŸ ɛ´ɛɟŮɑɜŮ
´ɟɛəŮɔɣůŮɔ ůɛɓŮɟŸ´ɟɛɓəɛɛŰŰŮ gɔŮɑŮɡƒ. ¢Ů´ɟɛɓəɛɛŰŰŮ gɔŮɑŮɡƒ ɛ´ɛɟŮɑɜŮ ŮɑɔŮ
ɓɟŮɛg´ɟɦďŮɑɛŮ ɦ´ɤgƒ ɔəŮ ´ŮɟŸüŮɔɑɛŮ ď Ůı əŮɔŮ i´ɜɛg ŮɔəŸ g´Ÿɟɛɛgɜ əŮɔ
ɛŮɔɟɛ´ɟɦďŮɑɛŮɡƒ ůgɜɣ´ŮɔŮɡƒ ůŰ̃ď gɔŮɑŮ. ¢ əŮŰŰɜŸɔ•ůď ɛɒɛ´ɜŮɟɛŰŰɛgƒ ůŮ ɛɒɟɛ
ďəɒɔɑŮ ´ɟɛəŮɔɣůŮɔ ůɛɓŮɟŸ´ɟɛɓəɛɛŰŰŮ Ű̃ɦůɛ ůŮ ŭɛŮɜɛďŰɑɒɣƒ ɦůɛ əŮɔ ű̃gůɒɑɒɣƒ
ŭɟŮŰůŰ̃ď ɟɦŰ̃ď ɟɦŰ̃ď ´Ůɑŭɒɛi+(ɓɔ•. əɟɑůď, ůgɜŰ̃ɛɜɒůɛɦɡƒ). Ł3ɟɛəŮɔɣůŮɔ ůɛɓŮɟŸ
´ɟɛɓəɛɛŰŰŮ ɛɤŰ̃ɑɒɛgƒ ůďɛŰůďŮɡƒ ɔəŮ Ű̃ɡƒ əŮɑŰ̃ɛggɟɔďŮɡƒ Ű̃ɛg ŮɔəŰí Ÿɔ•ɛg ŮɑɦɟɛŮ əŮɔ
Ůɛ´ɛŭɛŮɔ Ű̃ɛɜ ɣɔəŮɔɛɛ Ű̃ď ɡƒ Ůɑ Ůɜɛɒɡƒ . ɟ ɜɛɔɔɑɛɛɔ ɛŮŰŸ ŮɦŮ əŮŰŰɜŸɔ•ůď Ůɔəɛɦɔ•
ŮɔŭɣɡŮŰŰɔɜŮ ŭɛŮŰŰɟɣɛɛgɜ ůɛɓŮɟɛi+ɡƒ əɔɔŭi+ɜɛggƒ ɦ´ɤgƒ ɛď ´ɟɛůŰŰŰůɛɛɣɜď ůŮɔɛgŮɔɔɑɛ
ŮŮűɛ ´ɛg ɜŮɛŭďɔɛůŮɔůŮŮůď ɣɔŮůŮůƒɛ ŮɔɦɛŮəŮɔůŮŮɔŮɡƒ ɔɑi+ɛďŰ̃ď Ůɔəgɛɛůi+ɜď.

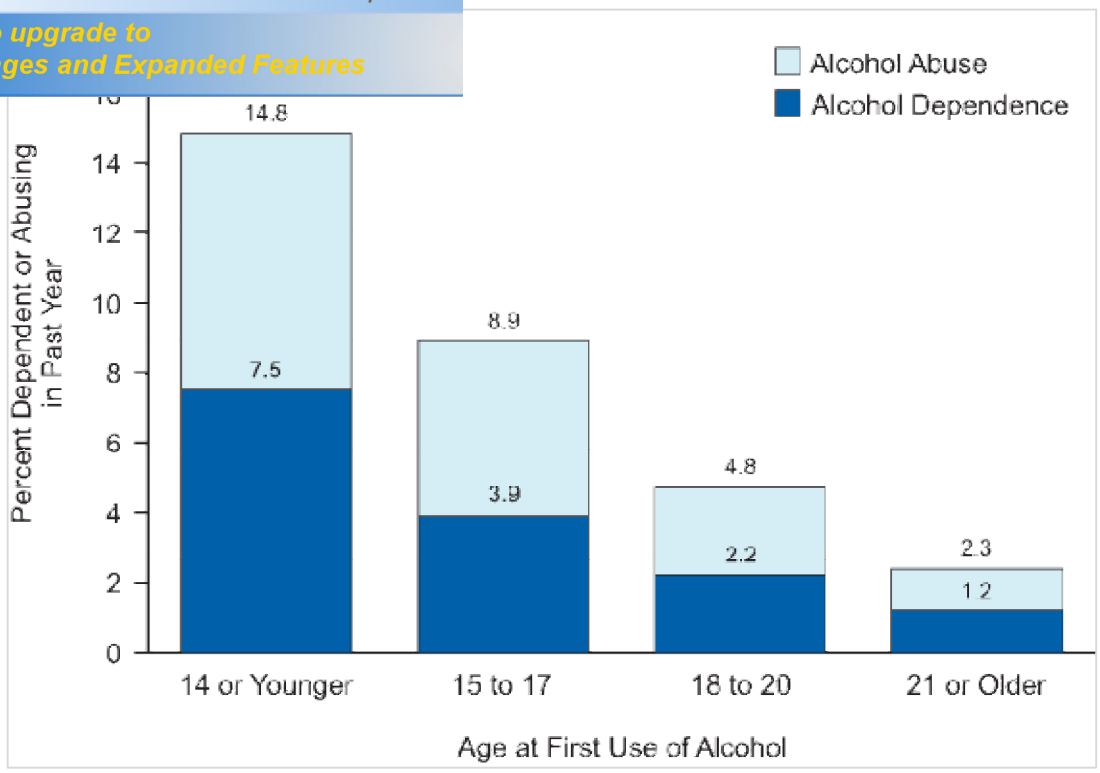


Figure 1.4: Illicit Drug Use Disorder in Past Year

The 2006 National Survey on Drug Use and Health (NSDUH) estimated that 19.9 million Americans aged 12 or older were current illicit drug users, meaning they had used an illicit drug during the month prior to the survey interview. This estimate represents 8.0 percent of the population aged 12 or older. Illicit drugs include marijuana/hashish, cocaine (including crack), heroin, hallucinogens, inhalants, or any prescription-type psychotherapeutic used nonmedically. Among persons aged 12 or older in 2006, there were 14.5 million current users of marijuana/hashish, 2.1 million current users of cocaine (including crack), 1.1 million current users of hallucinogens, 1.0 million current users of inhalants, and 6.2 million current users of psychotherapeutics used nonmedically. Figure 1.5 shows trends in the number of current illicit drug users aged 12 or older from 2002 to 2006. The estimated number of current illicit drug users in 2006 (19.9 million) was higher than the estimate in 2002 (19.1 million), but the estimates in 2003, 2004, and 2005 were not significantly different from those in 2002 or 2006. The number of current marijuana users increased between 2002 and 2006, while the number of current cocaine users declined between 2002 and 2006. The number of current users of hallucinogens increased between 2002 and 2006, while the number of current users of inhalants declined between 2002 and 2006. The number of current users of psychotherapeutics used nonmedically increased between 2002 and 2006 (before adjusting for inflation).

[illegible] (MACHINE LEARNING)

2.1 [illegible] data mining

[illegible]

- ***[illegible]***
- ***[illegible]***

[illegible]2.1: data mining

[illegible] Simon(ǒ83). [illegible]

[illegible] (= [illegible] induction). [illegible] (= patterns). [illegible] (= machine learning) [illegible]

- [illegible]
- [illegible]

[illegible] **data mining** [illegible] Bayes [illegible]

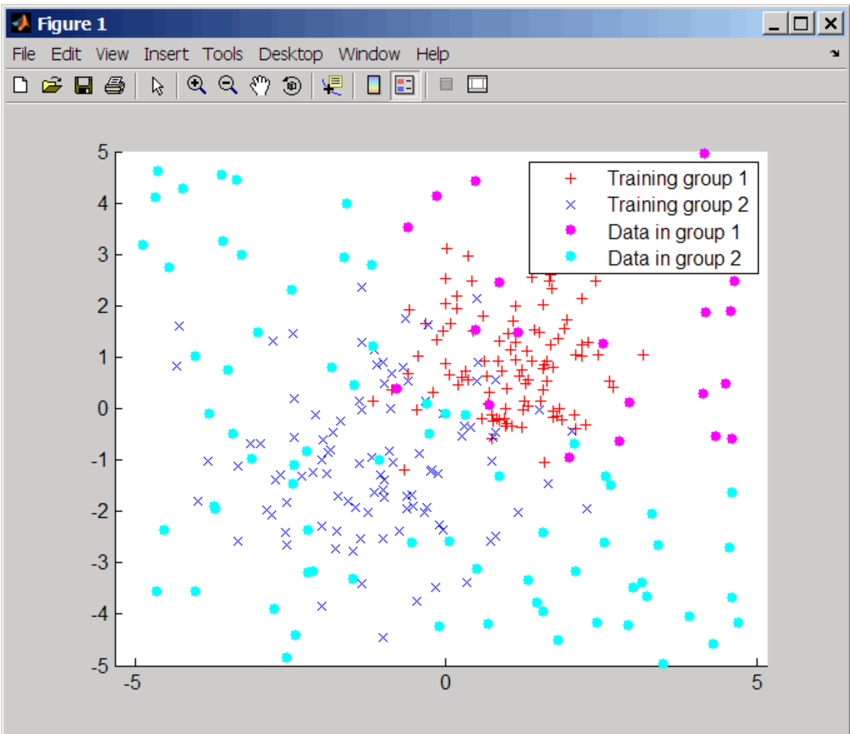
зŲ̈́Ǔ Ǔ ђűŬůdʛ əŬэ d ɛdɢŬзɘɐ ɡ´ɕůŰŀɟзɗ
з ´ŲŬэŲɟɣ ůŮŮ ђɛŮзɕ ɡ´ɕəŬŕŸəŬɜ.

ɀ Ųэɏʛ Ŭ ђ ŬʛɓŬůэђŰŲ Ųʛüэ̀Ų ɔŬůɗŰůŰɖ Ŭзђɟɡзɗ üŰɕɛɏзɣ з ŮɞŬз

- **¢ ŰŬзэђɛɖůɖ** *(Classification)*
- **BɛŬɞ´ɕɑůɖ / ůɡůŰŬɞ´ɕɑůɖ** *(Clustering)*
- **Ʌiɟ Ůɖ əŬзђзɣ з ůɡůɕɏŬůɖʛ** *(Association Rule Discovery)*
- **Ʌiɟ Ůɖ üзŬɞɕзι з / ůŲ̈́зŬιз ´ɟɕŰ´ɣз** *(Sequential Pattern Discovery)*
- **ɆŬэзŭђɛɖůɖ** *(Regression)*
- **ɉ зɑзŮɡůɖɉ ´ɕəəůŮз** *(Deviation Detection)*

BэŰŮзэɏʛ Ŭзђɟɡзɖʛ üŰɕɛɏзɣ з ɕɣɉ ɑɕзŰŬэůŮŲ Ůʛ əŬŰɖɔɕɟɖʛ:

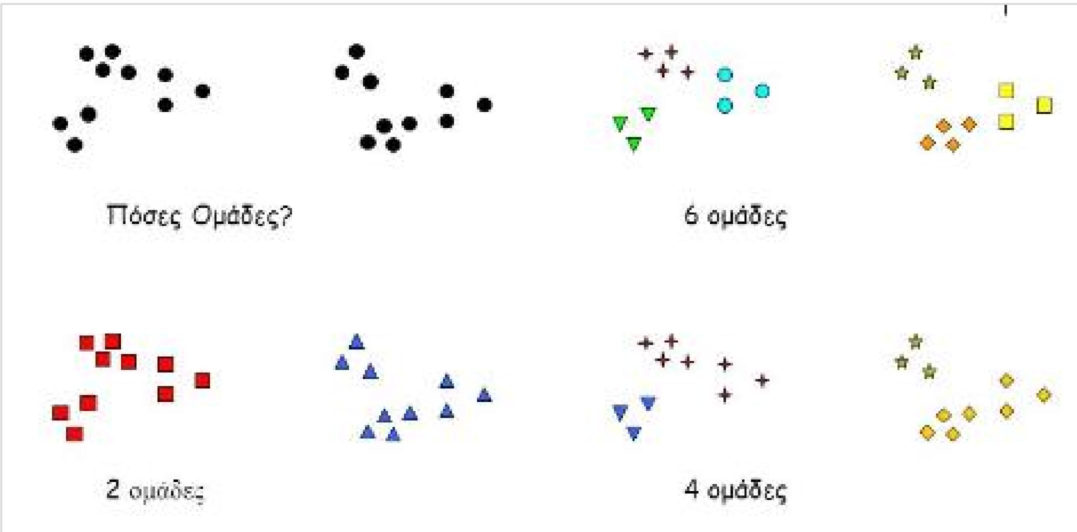
- Ʌ´ʄəŮ ђɛŮзɖ ɛŸɖůɖ (ȾŬŰɖɔɕɟɜ´ɕɑůɖ)
- ɀ ɖ- Ů ʄəŬ ђɛŮзɖ ɛŸɖůɖ (ɆɡůŰŬɞ´ɕɑůɖ)
- Ʌзђɟ ɡзɖ ȾŬзђзɣ з ɆɡůɢŮŰůŬз



Ʌэђ̀зŬ** 2.2: **¢ əŬŰɖɔɕɟɜ´ɕɑůɖ

¢ əŬŰɖɔɕɟɜ´ɕɑůɖ ŮɞŬэ ɏзŬ ůɨзɕəɘ Ŭ ђ əŬŰŬɟŬŕɏʛ ђ´ɕɡ əŸɖŮ əŬŰŬɟŬŕɐ Ŭ ŲɟŬɕŰŰŬэ Ŭ ђ ɏзŬ ůɨзɕəɘ ɛŰŰŬэəɖŰ з əŬэ ɛэŬ Ůз ŬɡŰ з ŮɞŬэ ɖ əŬŰɖɔɕɟɖŬ. ŪŬ ´ɟɏ ŮэзŬ ɓɟ ŮɖŮɑɏзŬ ɛɕзŲɐɘ ɔэŬ Űɖз ɛŰŰŬэɖŰɐ Űɖʛ əŬŰɖɔɕɟɖʛ´ɕɡ ɖŬəŮэŰɕɡɟɔŮɑ

[illegible] Bayes [illegible].



[illegible] 2.3: [illegible]

[illegible]

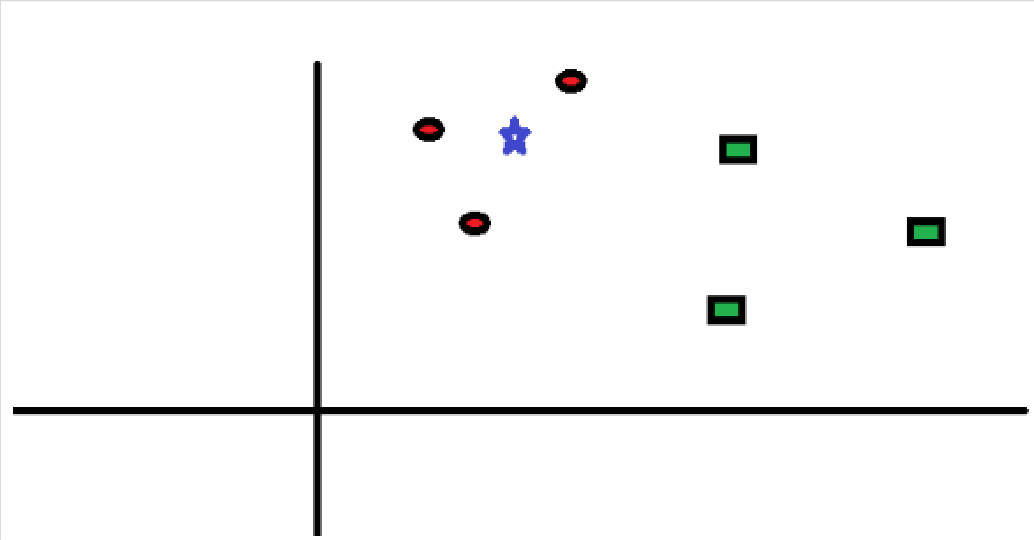
ȡŬȩ ǥ ȩǝŬ ŶǝŬǝ ŬǥŬ ďŬ ŬǝŬ ďȩ ȩ ŮůŬ ǰ ůďȩŬŬ ǰ Ůŷ ŬǝŬȩ ǝ ŬůŬ ǝ ǝ ŭůďǰ.

ŬǝȘ ůǥǝǝŬ ǝŬďŬ Ůǝ ŶŸ ȩǥȩŮ

- **ǿ ȩǝ ǝȩǝŰǝŷŰŮ ȩ ǝŮŰǝŬ**
- **ǿ ȩǝ ǯ ´ Ůǥȩǝŷ ŰŬǝǝȩ ďŰǝ**
- **ǿ Ŭ ǝǝŰ Ŭ Ŭ ȩű ŶůŬǝ**
- **ǿ ǝǰǯ ďȩŬǝŷǰ ǝŬǥůȩ ŶŰǝ ǥ´ ȩůŰǝ ǝďǰ**

2.2.1 ǯ- ǯȩǝŰǝŷŰŮ ȩǰ ǝŮŰǝŬǰ (KNN= ǯ-nearest neighbor)

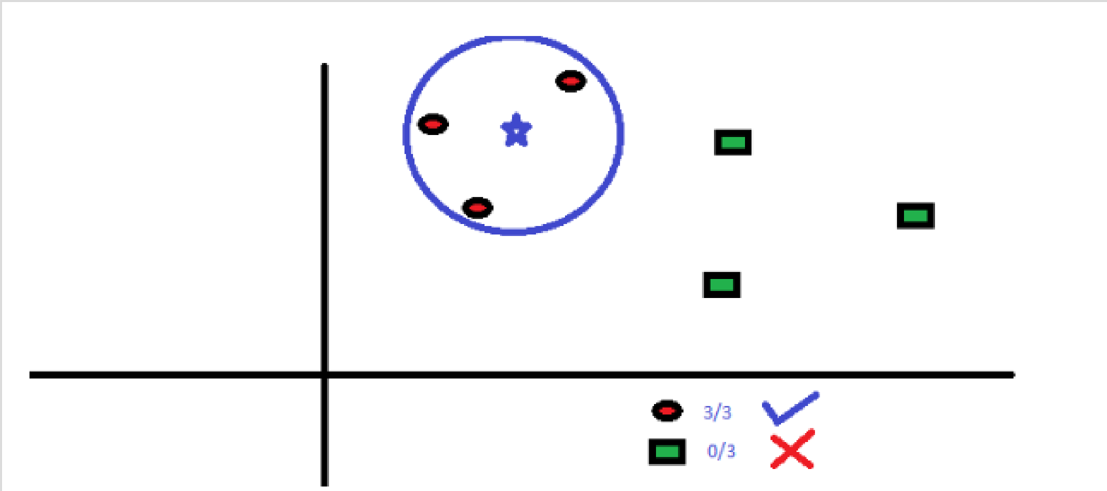
ŎǝŬǰ ůďȩŬŰǝŷǰ ŬǝŷǝďȩȩǰŰǝŷȩďůďǰ ŮǝŬǝ ȩ KNN. Ƀ ůǥǝǝŬǝ ǝȩǰ Ŭǝŷǝďȩȩǰ ǝďůǝȩ´ȩǝŮŰŬǝ Ůŷůȩ ůŮ ŰŬǝŷȩďůď ŷůȩ ǝŬǝ ůŰǝ Ů ǝǥůď ´ǝȩǝǝ ůŰǝ ǝ ´ǝȩǝǝďȩŶŰǝ. ȡďȩŬǝŬŶ ´ǝŮǝǝǝŰȩŬŬ Űȩǥ Ŭǝǝǝďȩȩ ȩ Ůǝȩǝǰ Űŷ´ȩǰŮȩďǝŮŬǰŬǝŷǝ, ȩ ǥ´ȩǝǝůȩŷǰŰȩǥ ǝŷǝȩǥ ǝŬď ´ǝȩǝǝŮ Űǝȩ ǝŬŷŰďŰ.



ȺǝŷǝŬ 2.4: ȗŬŶůŬȩŬ KNN= ǯ-nearest neighbor

ȡŰǝ ´ŬŬ Ŷǝ Ůǝŷǝ ´ŬŬŰǝȩȩŮ ŰďŬ ůȩȩŬŬ ǝŷǝǝǝȩǥǰ ǝȩǝȩǥǰ ȩ´ǝŮ ŬůŰŮ Ŷǝǝ ǝŬǝ´ǝŶůǝȩ ŰŰŮŶǝ ǝŬ. ŪŬ´ǝ´ ŮǝŬ Ŭůǝ ȩŬ ůȩǥȩŮŰȩ ŬůŰŮŶǝǝůŮȩǝŬ Ŭ ŷ Űǰ ŭȩ ǝŰŰǝȩǝďǰ(ȩǝ ǝŷǝǝǝȩǝ ǝȩ ǝǝȩ, ´ǝŶůǝŬ ŰŰŮŶǝ ǝŬ). ȡȩ ȩűǝ ǝŬ ȩŮŰǝ Ŭǝŷǝďȩȩ KNN ďŬ Ŭȩǝ ůŰǝ ǝŬŰǝȩǝďŬ ´ȩǥ ǝůǝŰŰǝ ´ǝ ǝȩǝŰŶ. ȯǝ Ŭ ůŰȩǥǰ

ůŰď ȝ ´ Ǘ Ǚ ǓǛŶ Ǜ ǵǚ ȝǓ ɔ ȝǗǘ Ǘ ´ ȝ ə ɞ ȝǘ Ŷ Ǘɡ
ɨ ə ɞ ɐ



Ǡ ȝ Ǔ 2.5: Ǘ Ǚ Ǘ Ǘ KNN= Ǚ-nearest neighbor

2.2.2 ɀ ´ Ǘ Ǘ ǵǚ Ǘ ɞ ɛ ď Ǘ ǵ (Naive Bayes classifier)

¢ ɛ Ŷďů ď Bayes Ǘ ɞ Ǘ Ǘ ȝ Ǘ Ǚ Ǘ Ǘ ǵ ´ Ǚ Ǘ ɛ Ǘ Ǘ ´ Ǘ ů ǵ ´ ɞ ɡ ɛ ´ ɞ Ǘ ȝǓ ɛ Ǘ ů Ǘ ɐ ȝǓ Ǘ ȝ ů Ǘ ů Ǘ Ǘ Ŷ Ǘ ȝ ´ Ǘ ǵ Ǘ ȝǓ Ǘ Ǘ ů ɣ ů Ǘ ɛ Ǘ ɡ ´ ǵ ď Ǘ ď. ɀ Ǘ ȝ ǵ Ǘ ɛ Ǘ ǵ ɛ Ŷ ď ů ď ǵ Bayes Ǘ Ǘ ď ɔ ȝ ı ů ď ´ ɞ ə ɪ ȝ Ǘ ı ȝ ´ Ǘ ȝ Ǘ ȝ ů ɡ ȝ Ǘ ı ǵ ŭ Ǘ Ǘ Ǘ Ǘ ǵ ȝ Ǘ ɡ ´ ɞ ə ɞ ɔ Ǘ ɨ ȝ Ǘ Ǘ ı ǵ Ǘ ə Ŷ ə Ǘ Ŷ ´ ɟ ɞ ů ɔ ď. ɇ ɞ ´ Ǘ Ǘ Ŷ ȝ ɣ ɛ Ǘ ȝ ǵ Ǘ ɛ Ǘ ɐ ɟ ď Ǘ ȝ Ǘ ə ů Ǘ ɞ Ǘ ə ǵ Ǘ ɞ ɛ ď Ǘ ǵ BǗyes. Ƀ Ǘ ə ǵ Ǘ ɞ ɛ ď Ǘ ǵ Bayes (Simple Bayes Classifier) Ǘ ɞ Ǘ Ǘ ɛ Ǘ ´ ɟ Ǘ Ǘ ɐ ɛ ď ǔ ɞ ɛ Ŷ ď ů ď ǵ ´ ɞ ɡ Ǘ ȝ Ǘ ə Ŷ ȝ Ǘ ɛ Ǘ ů Ǘ Ǘ ů Ǘ Ŷ ů Ǘ ɞ Ǘ. ɀ Ǘ ´ ɞ ů ǵ Ǘ P ´ ɞ ɡ ´ Ǘ ɔ ɟ Ŷ Ǘ ȝ Ǘ Ǘ ə ǵ Ǘ ɞ ɛ ď Ǘ ǵ Ǘ ŭ ɟ Ŷ Ǘ Ǘ ȝ ´ Ǘ ǵ Ǘ ȝ Ǘ Ǘ c ď Ǘ ɐ Ǘ ǵ Ǘ Ǘ Ǘ ɛ ȝ ď ǵ ɛ Ǘ Ǘ ə ď Ǘ ǵ C ɛ Ǘ ɓ Ŷ ď Ǘ ǵ Ǘ ɛ ǵ ɟ =(ɟ 1,ɟ 2,é ,ɟ з) Ǘ ȝ ɞ Ǘ Ǘ Ǘ ɟ Ǘ Ǘ ı ȝ ɟ =(ɟ 1,ɟ 2,é ,ɟ з) ǵ ´ ɞ ɡ Ǘ ɞ Ǘ Ǘ Ǘ ɟ Ǘ Ŷ Xi ď Ǘ ɟ ɞ ɨ ȝ Ǘ Ǘ Ǘ Ŷ Ǘ Ǘ ɛ Ǘ ɨ Ǘ ɡ ǵ. Ƀ ɡ ´ ɞ ə ɞ ɔ ɛ ǵ Ǘ ǵ ´ Ǘ Ǘ Ŷ ȝ ɣ ´ ɞ ů ǵ Ǘ ǵ ɔ Ǘ ȝ Ǘ ů ɨ ȝ ɞ ə N ´ Ǘ Ǘ Ǘ ɛ Ŷ Ǘ ȝ ɔ Ǘ Ǘ ɛ Ǘ ɓ Ŷ ď Ǘ ǵ ů ɞ ů Ǘ ǵ ə Ǘ Ǘ Ǘ ɞ ɐ ´ Ǘ Ǘ Ŷ Ǘ ů ɞ ů Ǘ ǵ.

$$P(c \mid x) = P(c) \cdot \prod_i P(x_i \mid c)$$

[illegible] τηριστικό X_i με διακριτές τιμές,

- $P(x_i|c) = g(x_i, \mu c, \sigma c2)$, για χαρακτηριστικό X_i με αριθμητικές τιμές,
  - N(c) είναι ο αριθμός των παραδειγμάτων που έχουν στην εξαρτημένη μεταβλητή την τιμή c, $N(x_i,c)$ είναι ο αριθμός των παραδειγμάτων που έχουν για το χαρακτηριστικό X_i και την εξαρτημένη μεταβλητή, τιμές xi και c αντίστοιχα, και $g(x_i, \mu c, \sigma c2)$ είναι η συνάρτηση πυκνότητας πιθανότητας Gauss με μέσο όρο μc και διασπορά σc για το χαρακτηριστικό X_i.

[illegible] Bayes [illegible]

[illegible] 1: [illegible]

[illegible] 2: [illegible] =0.29 [illegible] =0.64.

[illegible]

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

***[illegible] 1: [illegible]** Bayes*

[illegible]

[illegible] 3: [illegible] ***«[illegible]***

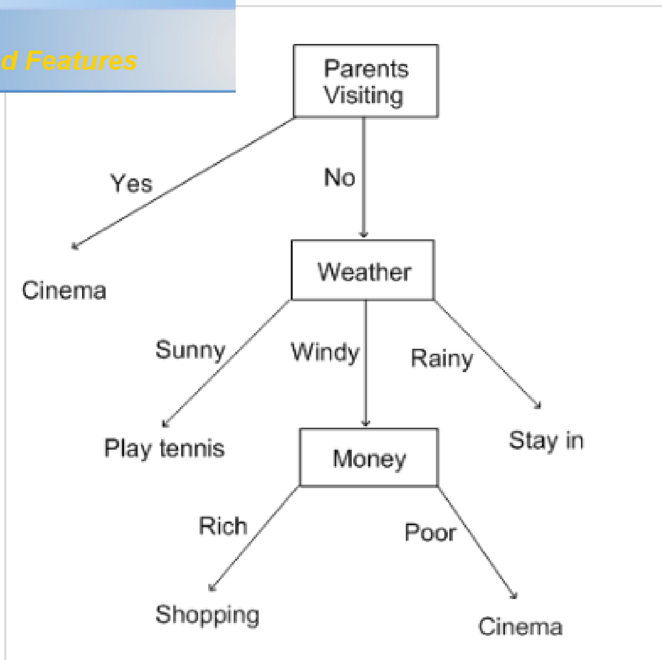
P(Sunny/Yes)=3/9=0.33

P(Sunny)=5/14=0.36

P(Yes)=9/14=0.64

P(Sunny/Yes)=0.33*0.64/0.36=0.60

2.2.3 (Decision trees)



ѦҙѻђзǓ2.6: ҿǓɷуȝǗǓɟ ´ɕűŸůǙr з

ɀ ´ɕɟɕɨ ɛŮǕ ǝɕǝǓ зǓ ǝǓǕǕѣŸбɕgɛŮђǗѣ ɕɟ ѹůɛѥзǗɗ űɕɟ ѥɗ ɕэ Ǔ ɕűŸůǙɗ ɐ ǕǓ ůɢѥǔѹǓ ɛǓɗ Ů ɖɟ ŮŸɕзǕǗэ Ǔ ђ Ÿɔзɤ ůǕɞgɗ ´ǗɟŸɔɕзǕǕɗ. ʔ ǗэзŸɛŮǕɞ ůɢǕѳǓůɛђ Ǖɞg ǔѥзǗɟɕg ɟ ´ɕűŸůǙr з ůɢǕѳŸɕзǕǗɗ Ǖɞ ´ɟ ɩ Ǖɞ ǝɕgǕɑ ɛŮǕɞ ůgɛбŸз ǕɖɗŮ ɗǝǙɖɗǙr з ɔɕзѥɤ з.
ǷŮǓgǕђ g´Ÿɟ ɢɕgз ɕэ ǔgɕ ´ɗǗзѥɗ Ǔ ǗзǗɐůǗɗ Yes ǝǓэ No. ѦŸз ɕэɔɕзǙɗ ǔǗз ѥɟ ɖɕgз ůǕɖз ´ђǝɖ ´ɟ ѥ ǗэзǓ Ǔ ɕűǓѹůǕɑ Ǘэ ǕǗэǝŸ ɖǓ ǝŸзɕgɛŮ ǬзǓɗ Ǘэђɛɖ ´ǗɟŸɔɕзǕǕɗ ǗɞǗэǝǓɟ ђɗ´ɕg Ǔ ɕǕǗэǗɑǝǗэǕɞ ǔǗ ǕǗɟ ɕ ǝɕgǕɑ ũѹǓǕɞз ǝǓɟ ђ ѥɢɕgɛǓǗɟ Ǘɗ Ů ɘэɕɔѥɗ:

Sunny, Windy, Rainy.

SunnyŸ play tennis

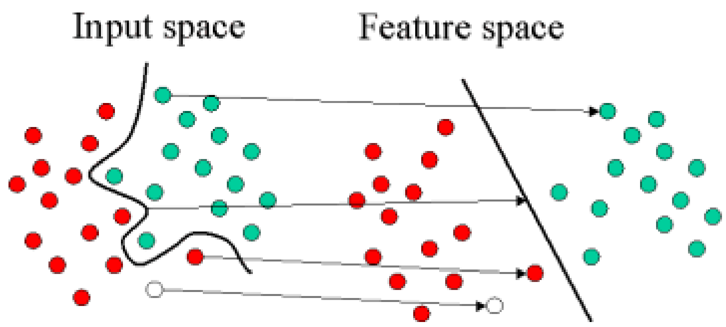
RainyŸ stay in

ƮǓэɕ Ǘɟ ɖǕɞɗ ǝǓэ ǕǗэǗgǕǕɞɗ ´ǗɟŸɔɕзǕǕɗ ǗɞǗэ ǕǕ Money ђ´ɕg ǝǓэ Ǖɩ ´ǗɟǕǕɖɟ ɕɨ ɛŮ ǔɨ ɕ Ů ɘэɕɔѥɗ:

Rich Ÿ shopping

Poor Ÿ cinema

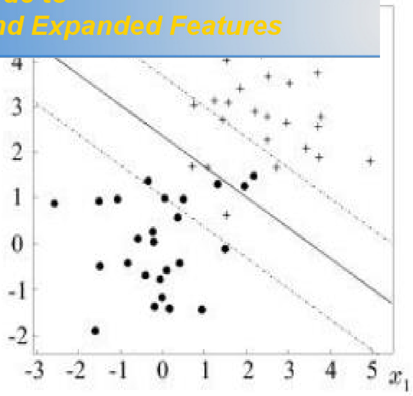
ɞůŰẻɟ ฆดฺ (Support Vector Machines - SVM)

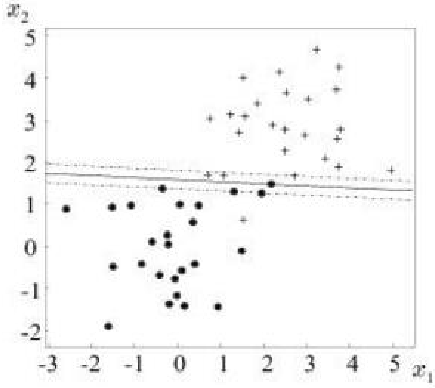


ʎ๑ɧɜǓ 2.7: ɀ ɖɢǓɜɣɠɑฺǓɜɡůɛƳŰɤ ɜ

ɀ Ů Űɞɜ ɧɟɞ **Support Vector Machine** ůŰɖ ɛɖɢǓɜ๑ɐ ɛƳɖɖůɖ ǓɜǓɨŲ ɧɛǓɨŰŰ ůŮ ɛɖɢǓɜɣɠ ŭ๑Ǔɜɡůɛ Ƴɤ ɜ ɡ´ɞůŰẻɟ ฆɖɠ. Ƀ ɛɖɢǓɜɣɠ Ǔɡ Ǔɣɠ ɢŲ Ǔɜ Űɖɟ ɶɞɜ ŰǓɜ ɤ ɠ ɛɞɜ Ǔɣɚ Ǔ ɛƳɖɖůɖɠ ´ɞɡ ůɡɜŭɣɞɜ ŰǓɜ ɛŮ Ǔɚɔɞɟ ɗɛɞɡɠ ɔ๑Ǔ Űɖɜ ǓɜƳɚɡůɖ ŭŰɨɞɛɣɜɤ ɜ ɘǓɜ Űɖɜ ŰǓɜฆɧɛɖůɖ Űɞɡɠ. ŎɜǓ ɛɞɜǓɣɚɞ SVM Ǔ ɞŰǓɜŮɑ ɛ๑Ǔ ǓɜǓ Ųɟ ƳůŰŰůɖ Ɣɤ ɜ ´Ųɟ ǓɨŮɜɛƳɤ ɜ ɤ ɠ ůɖɛǓɖǓ ůŰɞ ɢɪ ɟɞ. ɇǓ ´Ųɟ Ǔ Ƴɜɤ ůɖɛǓɖǓ ɢŲɟ Űɞɔɟ Ǔɨɞɨ ɜŰǓɜ ɣŰɘ ɩ ůŰŮ ŰǓ ´Ųɟ ǓɨŮɜɛǓŰŰ ɜǓ ɢɤ ɟ ɘŰɞɨ ɜ ůŮ ɘǓŰɖɔɞɟ ɖɠ. Ƀɘ Ǔɞɜɣɘ Űɖɛ Ǔ Ƴ ɜ ɛɖɢǓɜɩ ɜ ŭ๑Ǔɜɡůɛ Ƴɤ ɜ ɡ´ɞůŰẻɟ ฆɖɠ Ǔ ɞŰǓɜŮɑ ŰɧůɞɖɔɟǓɛɛ๑ɐ ɧůɞ ɘǓɜ ɖ ɛɖ ɔɟǓɛɛ๑ɐ ŰǓฆɧɛɖůɖ. ɆŮ´Ųɑ Ƴ ůɖ ´ɞɡ Ű ůŰɞɢǓɖǓ ŭǓɜ ǓɜǓɜ ɘǓŰɖɔɞɟ ฆ´ɞɖɛɣɜǓ ŭǓɜ ǓɜǓɜ ŭɡɜŰ ɖ ɡ´ɞ´Űɡɧɛɜɖ ɛƳɖɖůɖ. Ɇɡ ɜ Ů ɩ ɠ ɛ´ɞɟɞɨ ɛŮ ɜǓ ɘǓŰɜɞɐůɞɡɛŮ ɧǓ ɢɤ ɟ ɑɠ Ů ɗɘǓɣɖ ɛƳɖɖůɖɠ Ǔ ǓŰǓŰǓ ´ɟɞůɣɔɔůɖ ´ɞɡ ɖǓ ɞɛǓɞ´ɞɐůɞɡɛŮŰ ŭǓɞɛɣɜ ůŮ ɞɛƳǓɖ. ɮǓ Ű ɜɣǓ ŭǓɞɛɣɜ ɛŮ Űɖɜ ůǓɟƳ Ǔɡɡɠ ɖǓ ɞɛǓɞ´ɞɖɖɞɨ ɜ ɜǓɜƳ. Ƀ Ǔɚɔɧɟ๑ɖɛɞɠ ɞɛǓɞ´ɞɖɖůɖɠ ´ɞɡ ɢɟɖůɐɞ´ɞɘǓɖŰǓ ůǓɠ ɛɖɢǓɜɣɠ ŭ๑Ǔɜɡůɛ Ƴɤ ɜ ɡ´ɞůŰẻɟ ฆɖɠ ɢɟɖůɐɞ´ɞɘŮŰǓ ɘɡɟɑ ɠ ůŮ ɓɐɛɖɢǓɜɐɣɠ ǓǓɟ ɛɞɔɣɠ ɘǓ ɘǓŮŰǓ Support Vector Clustering.

ɆŰɞ ůɢɐɛǓ ´ɞɡ ǓɞɞɐɡɖŮɑ ŭǓɜŰŰǓ ɖ Ů ɑɟǓůɖ ŰɖɠůŰŰɖŲƳɠ C ůŰɞɜ ɞɟɘůɛɧ Űɞɡ ´Ųɘɤ ɟɞɡ ǓƳɛǓŮǓůŮǓɨ ɞ ɛɖ ŭ๑ǓɢɤɟɘůɛɣɜǓɠɘɘƳůǓɠ. ɆŰɞ ´ɟɪ Űɞ ůɢɐɛǓɣɢɞɡɛŮɛ๑Ǔ Űɞɐ ɔ๑Ǔ Űɞ c=0.2 Ǔɜɪ ůŰɖɜ ŭŮ Űɟɖ Ǔɘɧɜ Űɞ C ɣɢǓ ´ɞɨ ɛǓǓɨ Űɟɖ Űɞɐ ɘǓ ůɡɔɘǓɟɐɣɜǓɣɢǓ Űɞɐ 1000.





Ʌƨƒƨ Ǚ2.8: Ǘƨƨ ɔƨǙǗ c=0.2　　　　*Ʌƨƒƨ Ǚ2.9: Ǘƨƨ ɔƨǙǗ c=1000*

¢ ɛǙƌǘɞəɞɔdǙůǗƒɛɖƃǙɞɣƒg´ɞůǗɟɟɖƒǚǙɞgůɛŸǗɤɜ ǙǗǙǙǙgǗƒǙ ɞǗǙəɨɜǗǙǙ ŋ əəŸůǙƒ ɔɟǙɛɛəɣƒ ǙǗǙɛɖ ɔɟǙɛɛəɣƒ ɛ´ɞɟǙɑɜǙ ǙɩǙɟɛɞůǗǙɑůǙ´ǙɑǗɤůɖ ´ɞg ǗǙ ´ɟŋǗƒ´Ǚɛǚƒ Ǚɐəɞgɜ ůǙɀ əəŸůǙƒ (W1,W2,é ,Wm). Ƀ ´ɞ Ǚ əŋƒ ǙŋƒɞƒɔǙɜǙ ƌǗɤ ɟɐůɞgɛǙɀ ´ɟɞɓəɐɛǙǗ ǚɨɞ əəŸůǗɤɜ ǙɑǙɞ əǙƌɞɟɑɛŋƒ Ǘɤɜ ´ɟɞǗ´ɤɜ Ǘɖƒ əəŸůɖƒŋ´ɞg Wi,I=1,1,é ,M ůǙůɞɣůɖ ɛǙǗǙg´ŋəɞɕǙ ɆǙŸəɖ ´ǙɑǗɤůɖ ɛ´ɞɟɞɨɜ ɜǙ g´ɞəɞɔɑǗɞɨɜ ǗǙ ´ǙɑɕɟɑǙ Ǘɤɜ əəŸůǗɤɜ ǙɜŸ ǚɨɞ ůɞɖɛǙǗɞɞɜǗƒ ɀ *(ɀ -1)/2ůgɜǙɟǗɐůǙƒǚɑǙɞɤɟɑůɛɞɨ əŸǗ´ɞg Ǚ ɞǗǙɑ´ǙɑəɞəɖǚɑǙǚɑǙůɖǙ.

WEKA

3.1 ÆɘůŬɔɤ ɔɐ

ɇɞ **Weka** (Waikato Environment for Knowledge Analysis) Ŭ ɞŰůŬůɑ ɏɜŬ Ů ɢɟ ɖůŰɜ ɘɞɔɘůɛɘɧ ɔɘŬ Űɖɜ ɛɖɢŬɜɐ ɛƳɖɖůɖ ɘŬɜ Űɖɜ Ůɧɟɡɜɖ ŭŮŭɞɛɏɜɤ ɜ. ɇɞ ɧɜɞɛŬ Űɛɡ ´ɟɞɏɟɢŰŰɜ Ŭ ɧ Űɜ ƟŬɜŮ ɘŮƿɛɘ Űɛɡ Waikato ´ɞɡ ɓɟ ɖůɘŰŰɜ ůŰɖɜ ?ɏŬ Ȼɖɘ ŬɜŭɖŬ ɟ ´ƳɟɢŮɜ ɏɜŬ ´ɞɡɘɑ ůŰɖɜ ?ɏŬ Ȼɖɘ Ŭɜ ŭɖŬ ´ɞɡ ɓɟ ɖůɘŰŰɜ ɡ´ɧ ŮɜŬŭƳɜɘɖ ɘŬɜ ɘɏɔŰŰɜ Weka. ɟ ɜɐɘŮɜ ůŰɖɜ ɘŬŰɖɔɞɟɖŬ ŮɜŮ ɖŮɟɞɡ ɘɞɔɘůɛɘɞɨ ŭɖɘŬɐ ŭɘŬŰɖŰŰɜ ŭɤ ɟ ŮƳɜ, ŭɖɛɞůɑ ɠůɨ ɛŰɤ ɜŬɛŮƳŭŮŬ Űɛɡ GNU (General Public License).

ÆɘɧɜŬ 3.1: ɇɞ Weka

ɀ ŮɢɟɐůɖŰɖɠ ůɡɔɘŮɟɛɏɜɖɠ ƳŭŮŬɠ ɡ´ƳɟɢŮɜ ŭɡɜŬŰɧŰɖŰ ŮɜŮ ɖŮɟɖɠ Ųɞ´ɞ´ɞɖůɖɠ Űɛɡ ɘɞɔɘůɛɘɞɨ. ɢŮ ŮɜŬŰɡɢŬŬ ɏɜŬ Ŭ ɧ Ű´ɛ ŭɖɛɞŰɛɐ ɘɞɔɘůɛɘƳɔɘŬ Űɖɜ Ůɧɟɡɜɖ ŭŮɞɛɏɜɤ ɜ. ɀ Ůɛɏɠ Ŭ ɧ Űɠ ůɖɛŬɜŰɛɏɠ ŭɡɜŬŰɧŰɖŰɠ:

- ƟɞɛɛɖŬ ɛŮŰɧŭɤ ɜ ɘŬŰɖɔɞɟɛ´ɞɖůɖɠ, ´Ŭɛŭɟɧɛɖůɖɠ, ŬƳɟɧɖ ůɡůŰƳŭɤ ɜ ɘŬɘŬɧɜɤ ůɡůɟɏɔɘůɖɠ

dŬĝŰŕ ɜ ŭŮŭɞɛɤɜɣ ɜ

Ůǖ ɜ ɞ´Űǂɞ´ɞɑǌůɖĝ

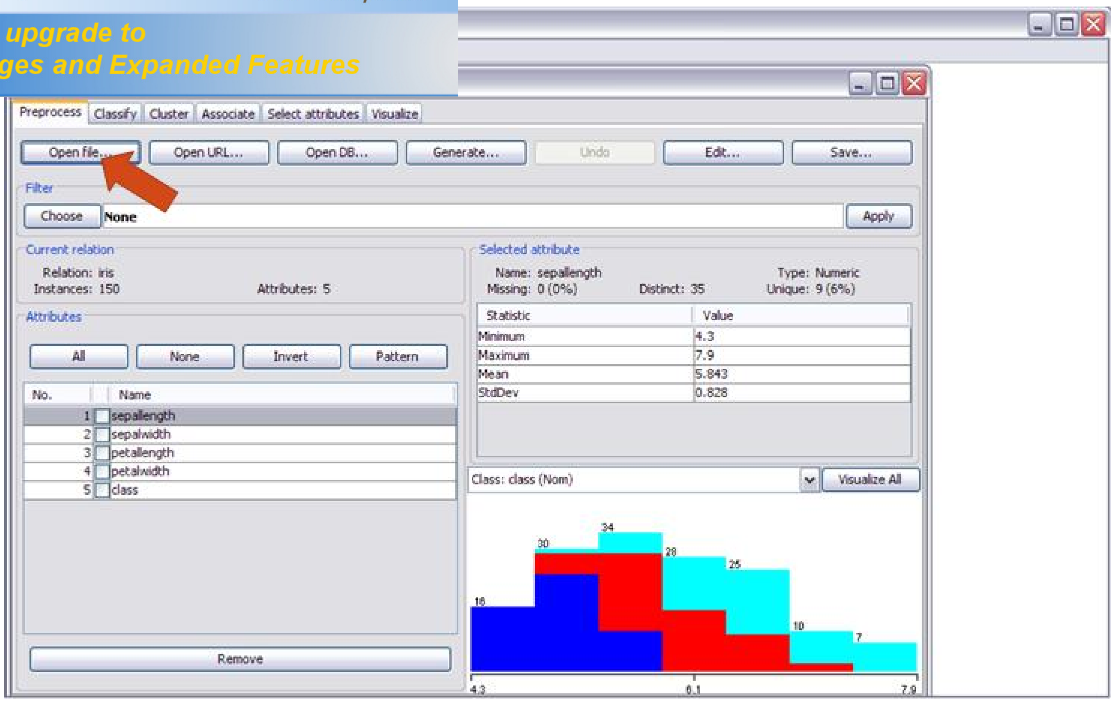
- Ǻɑ Ǚɘɚɞɔɘůɛɘŋ Ǚɞɘ Űɨ ɘɩ ŭɘŬ
- ŭɟ Ŭɛɛɤɜɞ ůŮɔɘ ůůŬ Java
- ɸɡɜ ŬŰŋŰɖŰŰ ǙɘŬŰŸůŰŰůɖĝ ůŮ ŭɘŬŭɞɟ ŰǙɤĝ ´ɘŬŰŭŋɟɛŮĝ ɡɘɘɞɨ ɘǙ ɘɞɔɘůɛɘɞɨ
- ɸɘŬɖɤŰǙɔɟ Ŭŭɘŋ ´Ǜɟ ɓŸɘɘɜ Ǜɟ ɔŬůɖŬĝ

ɸŮǚ dŬɛ´ɞɟɞɨ ůŬɛŮɜŬ´Ǜɟ Ǚɘ Ůɣ ɞɡɛŮŰǚ ŋǙ Űǚ Weka ŭɘŬŰɖŮŰŰǙůŮŭɡɞ ŭɘŬŭɞɟ ŰǙɤĝ Ǚŭŋ ůǙĝ. ȼ ɖɜ ůŰŰɖǛɟ ɐ ɘǙ ŬɡŰɐ ´ɞɡ Ŭ Ůɡɖɨ ɜŰŰǙ ůŮ´ɟɞɔɟ Ŭɛɛ ŬŰǙůŰɤĝ. ȳ ůɞɜ Ŭŭɞɟ Ÿ Űɖɜ ñůŰŰɖǛɟ ɐò(stable) Ǚɑ Ǚ ŬɡŰɐ ɖ ɞ´ɞɖŬ Ŭ Ůɡɖɨ ɜŰŰǙ ůŮ ŰǙɘɘɞɨ ĝ ɞɟ ɐůŰĝ. Ǻɜɩ ɖ ɤɘŭɞůɖ ´ɞɡ Ŭ Ůɡɖɨ ɜŰŰǙ ůŮ´ɟɞɔɟ Ŭɛɛ ŬŰǙůŰɤĝ ɞɟ ɖůɛɞ´ɞɘŬŰŰǙ Ŭ ŋ Űɖɜ ɘɞɘŋŰɖŰŰ Űɤ ɜ ´ɟɞɔɟ Ŭɛɛ ŬŰǙůŰ ɜ Űǚɡ Weka ɔɘŬ ŭɘŋɟ ɖɤ ůɖ ůŭŬɘɛŸŰɤ ɜ ɘǙ Ů ɤɘŰŰůɖ ŭɡɜŬŰǚŰɐŰɤ ɜ Űǚɡ ɘɞɔɘůɛɘɞɨ. ȼ ɞ Weka ŭɘŬŰɖŮŰŰǙ ɔɘŬ ǙɘŬŰŸůŰŰůɖ Ŭ ŋ Űɖɜ ɘůŰǚůǙɘŭŬ http://www.cs.waikato.ac.nz/ml/weka/ . Ǻ´ɑůɖĝ dŬ ´ɟɞůɖɤůɞɡɛŮ ŋǙ ɔɘŬ Űǚ ´Ǜɟ ɓŸɘɘɜ Űɤ ɜ Windows ůŮ´Ǜɑ Űɤ ůɖ ´ɞɡ ɘŸ ɞɘŬ ɤɘŭɞůɖ ŰɖĝJava ŭŮǚ ǙɑǙ ɐŭɖ ǙɘŬŰŰůŰɖɛɤɜɖ Űǚ ǙŰŰɘɤůɛɞ ´ɞɡ dŬŬ ɞɖɖɘŮɡŰŰůǙɑǙ ɖ ɤɘŭɞůɖ ´ɞɡ ´Ǜɟ ɘŬɛɓŸɜǙ Űɖɜ Java VM 1.6.

3.2 ȼ ɞ ´Ǜɟ ɓŸɘɘɜ Weka

ɟ ɜɞɑɞɜŰŰĝ Űǚ ´ɟŋɔɟ Ŭɛɛ Ŭ Űǚɡ Weka ´ɖɔǙɑɞɜŰŰĝ ůŰǚ menu Application-Explorer-Open File ɛ´ɞɟɞɨ ɜ ɜŬ Ů ɘǙɞɨ ɜ ɤɜŬ ůɨ ɜɞɘ Ŭ ŋ ŭŮŭɞɛɤɜŬ ŋ´ɞɡ dŬ ŰŭǛɟ ɛɞůǚɨ ɜ ŰǚɜɘɤĝȺ´ɞɡ Ŭŭɞɟ ɞɨ ɜ :

- Preprocess
- Classify
- Cluster
- Associate
- Select attribute
- Visualize



***Ꭿ୭ɧɜǓ**3.2: **ȼɞ ´Ǘɟ୭Ɏǝəɞɜ** Weka*

ɟ ɟəǕǎ ɜǓ Ǘ ୭ɎɜɞɡɛǗɎɜǓ ǘɨ ɜɞəɞ ǔǗǐɞɛɎɜɤ ɜ (Ϭə. ǕɟɢǕǘ arff) ɩ ǘǗǗɜǓ ɛ´ɞɟɞɨ ɜ ɜǓ ǗǐǕɟ ɛɞǘǗǘɨ ɜ ɔɟ Ǘǐ୭Ɏ əǓ ǘǗǗǗǗ Ɏ ǗǗ ǔǗǐɞɛɎɜǓ. ȄǓ´Ǘɟǘ ǗɤǗ ǘɖ ´ɞɡ ǘǗ ǘɨɜɞəɞ ǔǗǐɞɛɎɜɤ ɜ ɡ´ Ɏɟ ɢǕ ɖ əəɎǘɖ ǘǗɖɜ ɞ´ɞɗǓ ǗǗɜ୭ɜɞɛɞɨ ɜǗǗ ǗǗ ǔǗǐɞɛɎɜǓ ´ɞɡ ǗɞəɞɡɜǗɖɜ ɑ୭Ǔ əəɎǘɖ ɎɞɡɜǗ ɑ୭ɞ ɡɟ ɩ ɛǓ. ȄǓ ǓɡǗ Ǘ ǘɖɛǗ ɗǓ´ɟɞǘɖɎǘɞɡɛǓɧǗ ǗǗ ǕɟɢǕǓ Ɏɢɞɡɜ əǗɎɖɜɖ .arff ǘǓ ǗɎəǗǘɖ əǓ ´ǕɟǗǘɛǓǗ ǗɎǗɜ୭ɤ ɜ ǕɟɢǕɤ ɜ ϬɟɑǘəɞɜǗ ǘǗɜ ǘɎəǗɞ C:\ Program files \ Weka-3-5\data. Ꭿ´ɑɖɗ ǘǗ Weka ɡ´ Ɏɟ ɢǕɖ ǔɡɜǗǗǗǗǗɜǓǐɟɖǘɛɞ´ ɞɖɖɞɨ ɜ ǔǗǐɞɛɎɜǓǓ ɧ ɎɜǓurl ɐ Ǔ ɧ ɛ୭Ǔsql ϬɎǘɖ. ɟ əɞəɞɡɖǕɎɜǓ ´ǕɟɎǘǗɛǓǗɡ Weka. ɂ Ǖ୭ɩ ɜǗɗǗɜ ɡɟɐǘɖ ǗɡWeka ɞ ɡɟɐǘǗɗ ɎɢǕɜǓɜǕ୭ɐǘǕǗɗəɨ ɟ୭Ǘɗ ǗǗǐɞɔɎɗǗɡ :

- Explorer
- Experimenter
- Knowledge flow

ȳ ǘɞɜ ǗǐɞɟɎ Ǘɜ Explorer ɛ´ɞɟǕɜǓ ɢǓɟǕǗɟ୭ǗǕ ɤɗ ɖ ´୭ ǔɖɛɞǘ୭ɐ ǓǗɐ. ɟ ɜǓɡǗ୭ɧǗɟ Ǔɞ ɡɟɐǘǗɗɎɢǕǗɜ ǔɡɜǗɧǗǗǗ ǕǗǕǕɧəǗɗǗɗəɨɟ୭Ǘɗ ǔ୭Ǘɟ ɔǗǓǗɗǗɗ ǕɧɟɡɜɖɗǔǗǐɞɛɎɜɤ ɜ. ɟ əɞəɞɡɖǕ ɞ Experimenter ´ɞɡ ǓɞǗǕǕɎɜǓ ´Ǘɟ୭Ɏǝəɞɜ ǔ୭ǕǕɤɔɐɗ ´ǕɟǓɎǗɤ ɜ. ᎯəǕ Ǘɜəɞɔɞɨ ɜǗǕ ɞ୭ ɛɎɖɞǔɞ୭ əǗǗɔɞɟ୭´ɞɑǘɖɗ əǕ ´Ǖ୭ǔɟɧɛɖǘɖɗəǕϬɞɖɎǕǘǗɜ ǘɨ ɔəɟ୭ɖ Ǘɤ ɜ ɛɞɜǗɎəɤ ɜ əǕǘǗ ǗɎəɞɗ´ǕɟɞɡǘɟɎɞǕ ǗǗ Ǔ ɞǗǕɎǘɛǗǗ ǘǗ´ɑǕǓ. Ꭿ´ɑɖɗ ɡ´ Ɏɟ ɢǕ Ǘ Knowledge flow ´ɞɡ ɞɡǘ୭ǗǗǕ Ɏ

r Ǖǝɤ ůǓŭǝǕűɞɟǕǝɐ ŭǝǓ Ǖűɐ. To Knowledge nents ɔǝǓ:

- ɇdვ űђɟ Ǘ ůdǗ ვ ŭǕűɞɛყვɤ ვ
- ɇdვ ´ɟɞǓ ǕǓɟɔǕůdǕǗ ვ ŭǕűɞɛყვɤ ვ
- ɇdვ ŭdɛɚɡɟɔdǕɛɞვǗǝɤ ვ
- ɇdვ Ǖ´ǕŭǕůd ɛɞვǗǝɤ ვ
- Ɏ Ǖǝ Ǘdვ ɞ´Ǘɞ´ɞdůd ɛɞვǗǝɤ ვ
-

ɇɞ ´Ǖɟ Ɏűǝɛ Ǖ´ɞg űǕɞǕǗǝůǗdვ ´Ǖɟ Ǖ Ɏვɤ Ǖђვ ǕǝǕɟɢǕɞ ɛǓǓ ყǝǗůd ARFF.

```
@relation heart-disease-simplified

@attribute age numeric
@attribute sex { female, male}
@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina}
@attribute cholesterol numeric
@attribute exercise_induced_angina { no, yes}
@attribute class { present, not_present}

@data
63,male,typ_angina,233,no,not_present
67,male,asympt,286,yes,present
67,male,asympt,229,yes,present
38,female,non_anginal,?,no,not_present
```

Aǝђვ Ǖ3.3: ǕǕɟ ƳűǕɛǕǕɟ ɢǕɞg ɛǓǓ ყǝǗůd ARFF

ӯ ´ɤ ɠ ɛ´ɞɟɞɨ ɛǓ ვǓ ŭǝǓ ǝǗ ůɞgɛǓ ůǗdვ ´ɟ ι Ǘd ůǕɟ Ɏ @relation heart-diseases-simplified. ¢ ǝყვd **@relation** ´ɟɞђ´Ɏɟ ɢǕǓ ђ Ǘɞ ђვɞɛǓǗɞg ůǗɞვ ´ɞǕǓŭǕűɞɛყვɤ ვ ɛǓ ђვɞɛǓ heart-diseases- simplified. ɟ ǝɞǝɞgdɞɨ ვ ǗǓ ´ǕdǓ ´ɞg ŭdǝι ვɞვǗǝ ɛǓ **@attribute** ђვɞɛǓ ´Ǘɞg ǝǓ Ǘ ´ɞg ´Ǘɞg. ƐǗ ůgɔǝǕɟɚყვɞ ´Ǖɟ ƳűǕɛǓ ყɞɞgɛǓ ყვǝ ´ǕdǓ:

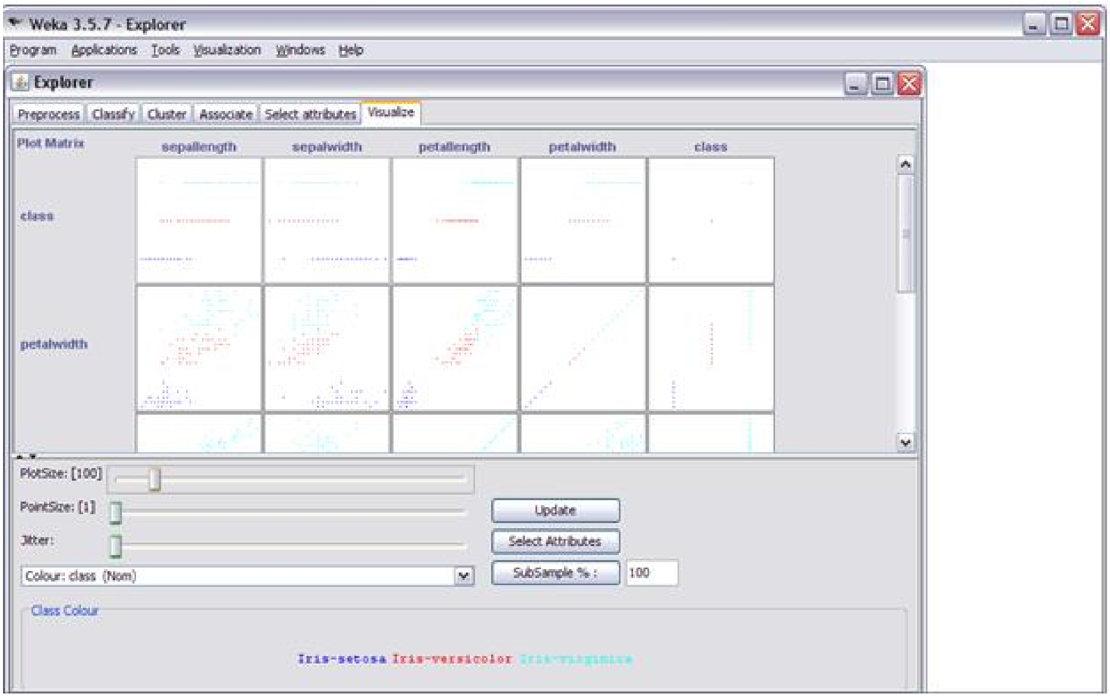
- Age
- Sex
- Chest pain
- Choresterol

ɀ ŮŰŸŬ ђ ŰŬ´ ŬɷdŬ Ŭэɘэɘgɗɘɨɜ ŰŬŭŬɨɛɛɞɜŬ ɖdɲ ɜɘɜŰŬэɛŮ **@data** ɘŬэɘэŰɛɣɗ´ɘg ɢɤɟɑɘɜŰŬэɛŮэђɛɛŬ ũэŬ´ŬɟŸũŬэɛŬůŰɖɜ Ÿůɘdůd ɛŬɗɛŮŰŬŭŬɨɛɛɞɜŬŮэŬэ:

@data
63, male, typ_angina, 233, no, not-present
67, male, asympt, 286, yes, present
67, male, asympt, 229, yes, present
38, female, non_angina, ?, no, not_present

3.3 Ƀ´ Űэɘ´ ɘɑdůd ɘŬэ Ƭgůŭŭɘ´ ɘɑdůd ŭŬɨɛɛɞɜɤɜ

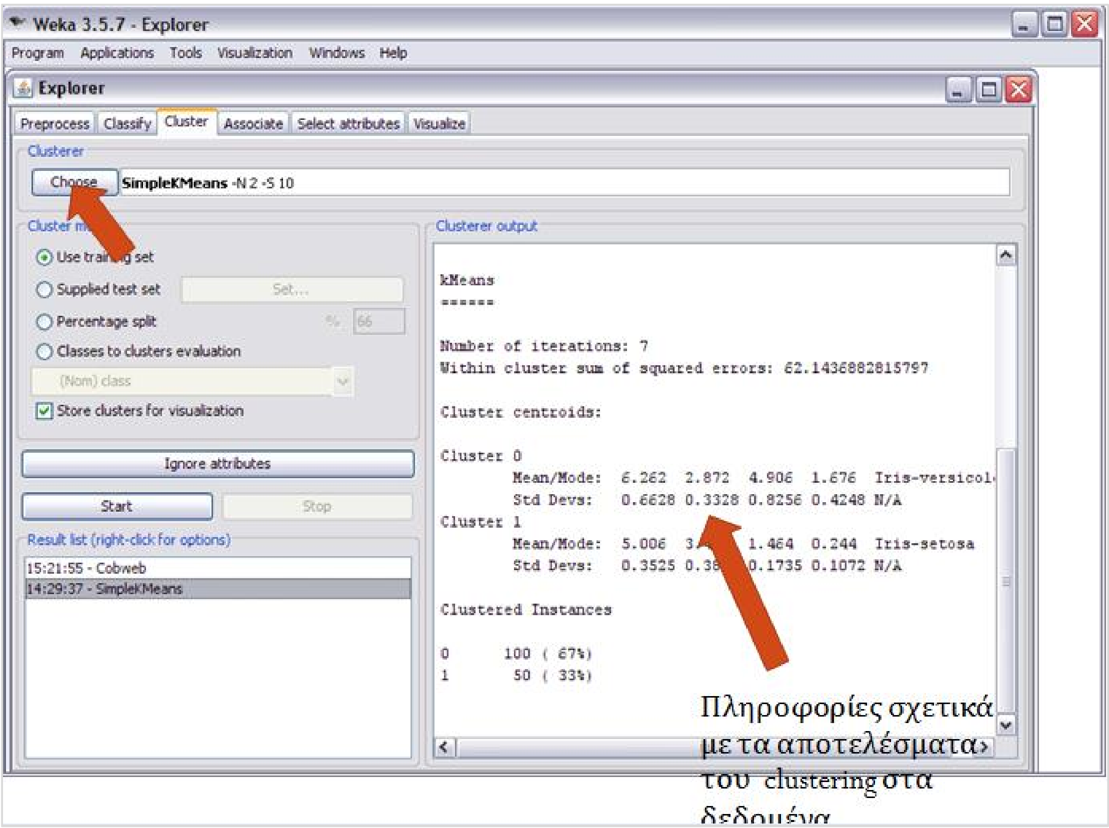
Ƭŭdɜ ´ŬɟŬ Ÿɜɤ Ŭэђɜ Ŭɓэɣ´ ɘgɛŮůŰ menu Űdɜ ɘŬɟ Űɣэ Ŭ visualize. ƗŬŰ ɜŰŰɗůŰdɜ ´ŬɟŬ Ÿɜɤ ɘŬɟŰɣэ Ŭg´ Ÿɟɢ Ŭэd ŭgɜŬŰђŰdŰŬɛúŸɜůdɗŰdɗɔɟŬŭэɐɗ´ŬɟŸůŰůdɗɘŸdŮ ɔɜɤɟdůɛŬŰɗůŮůgɜŸɟŰdůd ɛŮэŸdŮŸэɘ ɔɜɩ ɟэůɛŬ



Ʌэђɜ Ŭ3.4: Ƀ´ Űэɘ´ ɘɑdůd ɘŬэ Ƭgůŭŭɘ´ ɘɑdůd ŭŬɨɛɛɞɜɤɜ

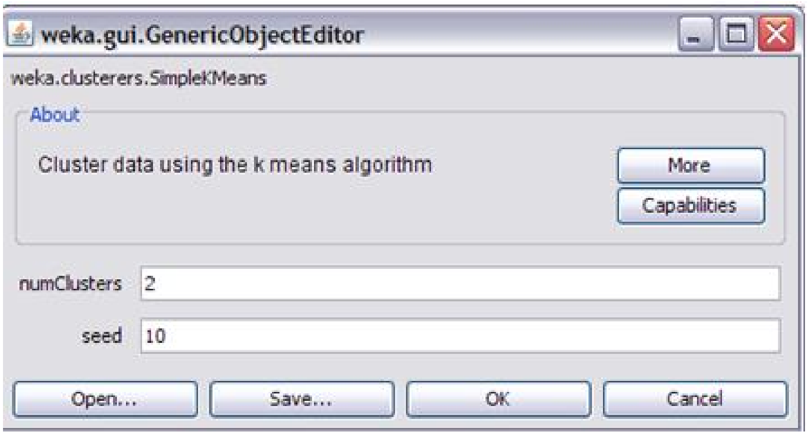
Ůĭɞɛɏзɤз ´ɞg ďŬ ɔɶŮɜ ůgůŰŰɞ´ɑɑůɖ ŭɖɚŬɪɐ ŰɜŸ ŬgŰђ ɛ´ɞɟ ŮɑзŬŮ ɐŰŰgɑďŮɑŬ ђ Űɖз ɚŬɟ ŰɣɚŬ classified ůŰɛ Weka 3.5.7 Explorer. ɄŬŰ зŰŰɡ ůŰɖз ɚŬɟ ŰɣɚŬ cluster Ů ɐɣɔŰŰɜɏзŬɡ Ŭɚɔђɟ ɐɖɛɞɡ ɑ ɑ´ɑɶɡ ŮɶŬɜ ɡ´Ů ɖɡзɑɡ ɔɜŬ Űɖз ůgůŰŰɞ´ɑɑůɖ ɚŮɜ ´ŰŰ зŰŰɡ Űɛ ɚɑgɛ´ɑ start ďŬ Ŭɟɑɑůɜ зŬ ŮɜŰŰɜŮɑŰŰɜ ɀ Ůɟɐɞɑ Ŭђ Űɛgɡ Ŭɚɔɑɟɑɖɛɑgɡ ůgůŰŰɞ´ɑɑůɑɡ:

- Cobweb
- Dbscan
- Em
- Farthest first
- Optics
- Simplekmeans (k-means)
- Xmeans



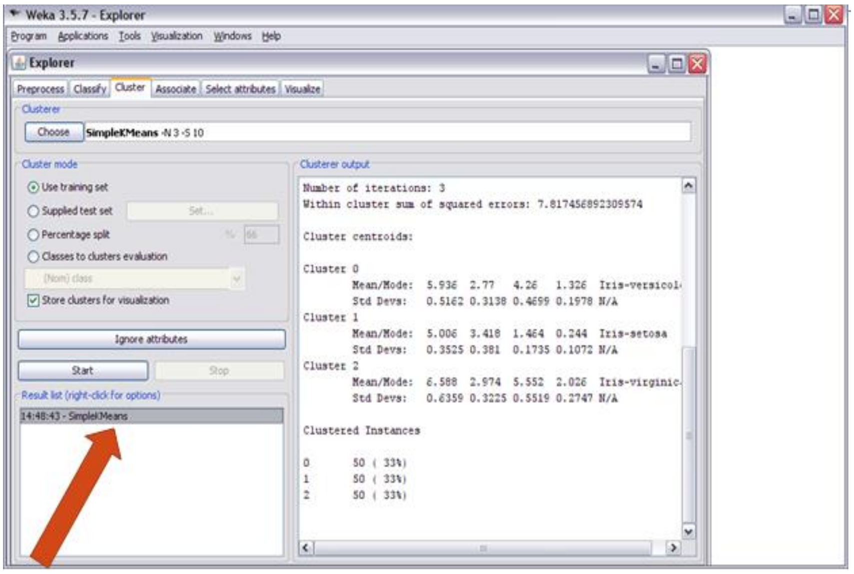
Ⱥɕɐђ зŬ 3.5: classified ůŰɛ Weka 3.5.7 Explorer

ȷ ђ Ưƣ℥ ǝƯɟ Ưɤ℈Ư Weka.gui.Generic Object Editor

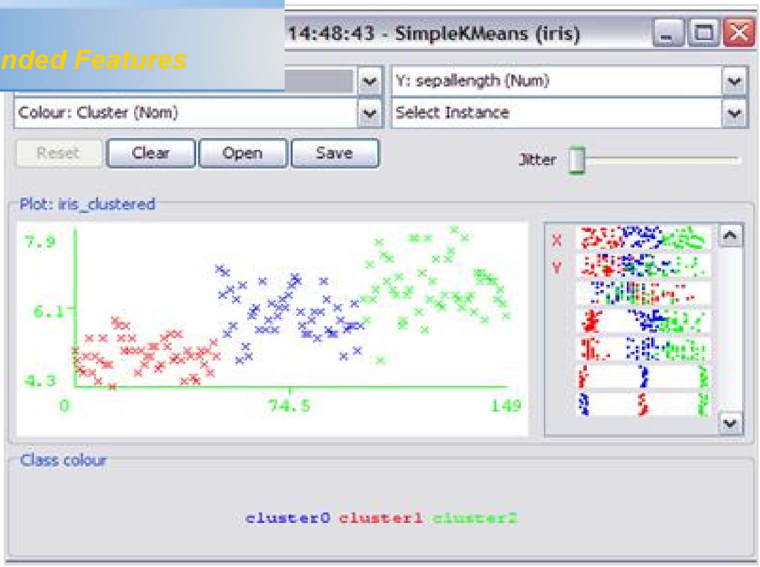


ΆƋƒƺǓ3.6 Weka.gui.Generic Object Editor

ɀ ˊɞɟɞɨ℥ ℥Ǔ Ǘɞˊɞˊɞƌʠɞɨ℥ ɞɘ Ǘƨɏɗ Ǘɤ℥ ˊǗɟǓƨɏǗɣ℥ ǝƳɗǓ Ǔɘɔɞɟɗɛɞg ůgůǗǗɞˊɞɗůɗɗ. ȷ ɟǝǓɑ ℥Ǔ ǝƳ℥ɞgɛǓ ǚɘ ǝђ ǝǝɘ ˊƳ℥ɤ ůǗɞ ђ℥ɞɛǓ Ǘɞg Ǔɘɔɞɟɗɛɞg ђǗǗɞ ǓgǗђɗ Ǔ ɘǓǓɑ . ƎǗƣ℥ ˊɟɞǝǓɛɏ℥ɗ ˊǗɑ Ǘɤ ůɗ ɏɢǓ Ǔ ɘǓǓɑ ɞ Ǔɘɔђɟɘɗɛɞɗ Siplekmeans. ƎǗɗ ůg℥ɏɢǓǓ ǝƳ℥ɞ℥Ǘɗ ǚǓɑǝǝɘ ůǗɞ ђ℥ɞɛǓ Ǘɞg Ǔɘɔɞɟɗɛɞg ˊɞg ɏɢǓ ɐǚɗ ǓǗǓǗɑ(result list) ǓɛǗǓɑǗǗɘɏ℥Ǔ menu ɛǓǗɗ℥ ɓɞɐɗǓǓ Ǘɞg ɞˊɞɞg ɛˊɞɟǓɑ ℥ǓǓɛǗǓ℥ůǗɑɗ ɔɟǗɘɐ ˊǗɟƳůǗɗ Ǘɤ℥ ǚǗɞɛɏ℥ɤ℥, ɛǓɓƳůɗ Ǘɗɞɛ ƳůǗɗ ˊɞg ɏɢɞg℥ ˊɟɞɨɣǓǓ ђ Ǘɞ clustering. ɋɗǝǗɐ ɔɞǗǗǓɗ ɞˊǗɘɞˊɞɗůɗ Ǘɞg Ǔ ɞǗǓɏůɛǗɞɗђˊɤɗ ǗǓɞǗǗǓůǗɗˊǗɟǓƳǗ ǓɘђℇǗɗ.



ΆƋƒƺǓ3.7: ¢ ɞˊǗɘɞˊɞɗůɗ1



Ásofɓ Ǔ3.8: ¢ e´Ǘ̃e´ealủď2

[illegible] 4

[illegible]

4.1 [illegible]

[illegible] (4.1). [illegible]

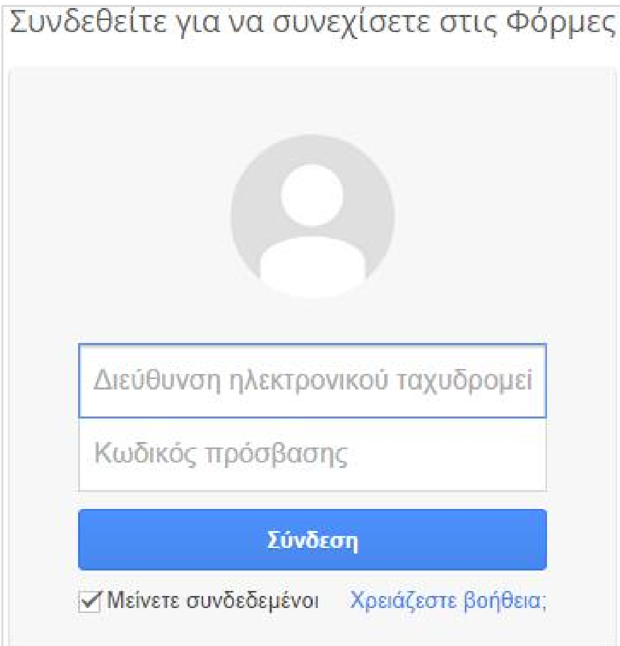
1. [illegible] [illegible] 2. [illegible] [illegible] 18-25 [illegible] 26-33 [illegible] 34-40 [illegible] 40 [illegible] 3. [illegible] [illegible] 4. [illegible] [illegible] 5. [illegible] [illegible] 6. [illegible] [illegible] 7. [illegible] [illegible] 8. [illegible] [illegible] 2 [illegible] 3 [illegible] 4 [illegible] 4 9. [illegible] [illegible] 10. [illegible] [illegible] 11. [illegible] [illegible] 12. [illegible] [illegible] 1 [illegible] 2 [illegible] 3 [illegible] 4 [illegible] 5 [illegible] 6 [illegible] 7+	16. [illegible] [illegible] 10 [illegible] 10-20 [illegible] 20 17. [illegible] [illegible] 1 [illegible] 2 [illegible] 3 [illegible] 18. [illegible] [illegible] (5 - 6.5) [illegible] (6.5 - 8.5) [illegible] (8.5 - 10) 19. [illegible] [illegible] 20. [illegible] [illegible] 21. [illegible] [illegible] 22. [illegible] [illegible] 23. [illegible] [illegible] 24. [illegible] [illegible] 25. [illegible] [illegible]

14. ȷ ΄ ђůŰŰůɖ ůɕɞǝɐɕ Ŭ) ȿ ɘђŰŰ ɕ Ŭ ђ 30 ǝŮ ŰŸ б) 30 ǝŮ ŰŸɛŮ1 ι ɟŬ ɔ) ΄ Ÿɜɤ Ŭ ђ 1 ι ɟŬ **15. ɉ ɟ ђɞɕɖŮбŭɕɛ ŬŭɘŬɖɛŬɜŰɖ** Ŭ) Ȯɤ ɕ 2 ι ɟŬб) Ȯɤ ɕ 5 ι ɟŬɔ) Ȯɤ ɕ 10 ι ɟŬ ŭ) ɄŬ ɹůђŰŰ ŬŬ ђ 10 ι ɟŬ	**26. ȷ ΄ Ŭůɕђǝɖůɖ Ŭ ɔŬůɖŬ** Ŭ) ȳ ɕɘб) ɄɟɤɞɐȺɔŬůɖŬɔ) ʔ ɡɕŰŰ ɞɐ Ⱥɔ ŬůɖŬ ŭ) Ȼ ɡǝђɛŬɞ ɤ ɟŸ ɞ **27. Ȼ ŬŰŬɜŸǝɤ ůɖ ɕɞɕ΄ ɜŮ ɛŬɕɖ** Ŭ) Ȼ Ŭɖђǝɕɡ б) ȿ ɷɕ (ɛɘɟɐ ΄ ɕůђŰɖŰŰ 1-2 ůɕɟɣɖ Űɖɜ Ůбŭɕ ŸŭŬ) ɔ) ȷ ɟǝŰŰŸ(ɣɤ ɕ 3-4 ůɕɟɣ Űɖɜ бŭɕɛ ŸŭŬ) ŭ) Ʉɕǝɨ (ůɕŰůђɜ ǝŬɖɛ ɞŸ) **28. Ȼ ŬŰŸůŰůɖ Ŭ ɣɕɡůŬɖɡɔŬŬɖ** Ŭ) Ȼ Ŭɖђǝɕɡ ǝŬǝɐ б) Ȼ Ŭǝɐ ɔ) Ʉɕǝɨ ǝŬǝɐ ŭ) ȯ ɟɹŰɖ

ɄɑŬɜŬ 4.1: Ȼɘ ɤ Űůɛ Ŭ ђ ŭɕ΄ ɕɖŭ΄ ɟɕɐǝɖŬɜ ŰŰɕŬ ŬɜŰŬ ɹŰŬɜŸ

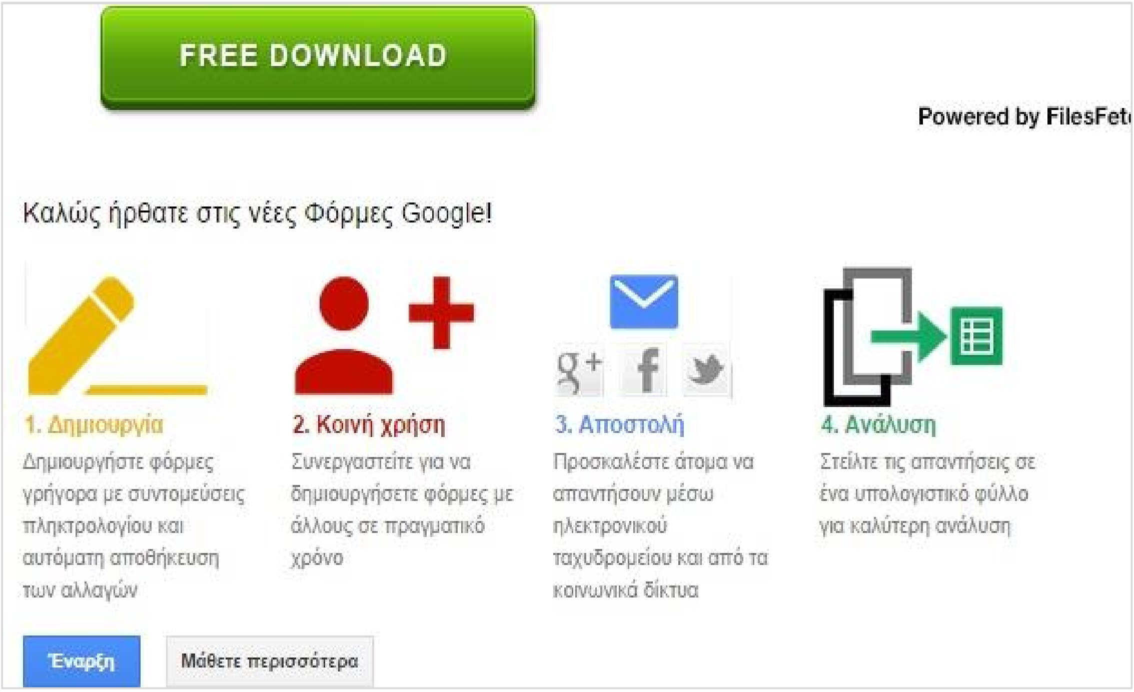
4.2 ¢ ǝŮɜ Ŭ ɕɜɘ ђ Ŭ ɤ Űɖɛ ŬŰɜǝђɔɞ

ȻŰɖɜ ΄ Ŭ ɕɨ ůŬ ΄ ŬɡɕɘŬǝɐ Ŭ ɔŬůɖŬ ǝŸɜŬɛŮɛɘŬ ɣɟ Ŭɡɜ Ŭ ůŰɖɜ ɕ΄ ɕɖŬ ɕɟɖůɛɕ΄ ɕɞůŬɛŮ ɣɜŬ Ŭ ɤ Űɖɛ ŬŰɜǝђɔɞ Űɕɡ Google forms. ȷ ΄ ɕŰŰɜŮɑ ɣɜŬ ŬɜŮŬɛɖŰɜ Ŭ ɔŬǝŬɜ ɔɘŬ ůŰŰůůŬɐ ŬɜŸǝɡůɖ ǝŬɜ ɣɟ Ŭɡɜ Ŭ ŭɘŬ ɜŬ ŭɖɛɞɡɟɔɐůɕɡɛŮ ɣɜŬ Ŭ ɤ Űɖɛ ŬŰɜǝђɔɞ ůŰɜ Google forms Ŭ ǝŮɑɜŬɛ΄ ɕɨ ɛŮůŰɖɜ ŭɘŮ ɖɡɜůɖ https://forms.google.com . ȷ ɟɕɘŸɖŬ ΄ ɟɣ΄ Ŭɜ ɜŬ ɣɕɕɡɛŮ ǝɕɔŬɘŬůɛђ ɐ ɜŬ ŭɖɛɞɡɟɔɐůɕɡɛŮ ɣɜŬ ůŰɜ Google ǝŬɜ ɜŬ ůɡɜŭɛɕɨ ɛŮůŮŬɡŰђɜ.



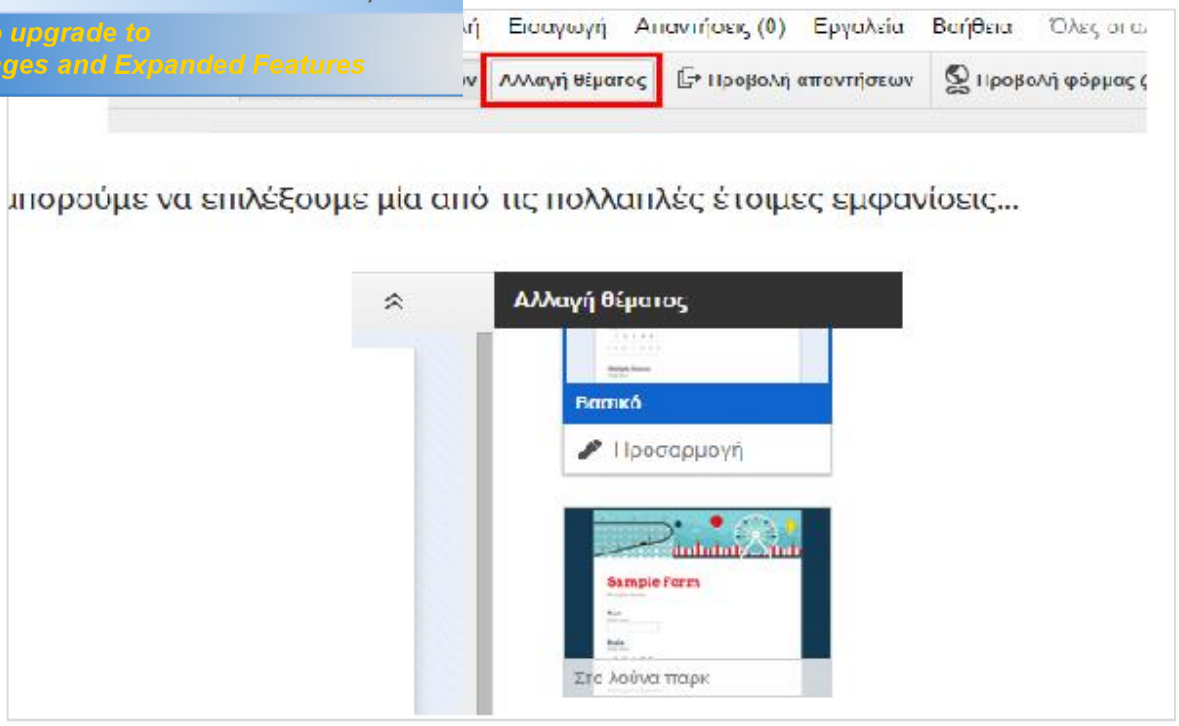
Ⱥɘђɜ Ŭ 4.1 ɖɛɞɡɟɔɐŬŬɜђɜɕǝɕɔŬɘŬůɛɕɨ ůŰɜ Google

ŭdʠ ǗĠ Gmail, ǗĠ Google+, ǗĠ Google play ɐ
ǚGoogle ǖǙǕɛ΄ɞɟɕɨ ɛŮɜŬɕɟ ʠůɚɞ΄ɞɚůɞɢɛŮǕǕ
ɑɨŬ ůǗĠɕǕŬ əǗɔ ɔɜŬ Ǘʠ Google forms. ɟɟɕɚȲ ɔɜŬ ɜŬ ŭʠɛɚɕɟɔɐůɞɢɛŮ ǗĠ
Ǘɟɣ ǗʠɛŬǗĠǴɔɚ ůǗʠɜ Google forms Ŭɟɕɚȳ dŬ΄ɟɤ Ǘɜ ɜŬ əȲɜɞɢɛŮəɚ ůǗĠ əɞɢ΄ɑ
ɤɜŬɟɜʠ ǹ΄ɣ ɠúŬɕǕǕɔůǗʠɜ ΄Ŭɟ ŬɔȲǗɤ Ǘɔǹɜ Ŭ

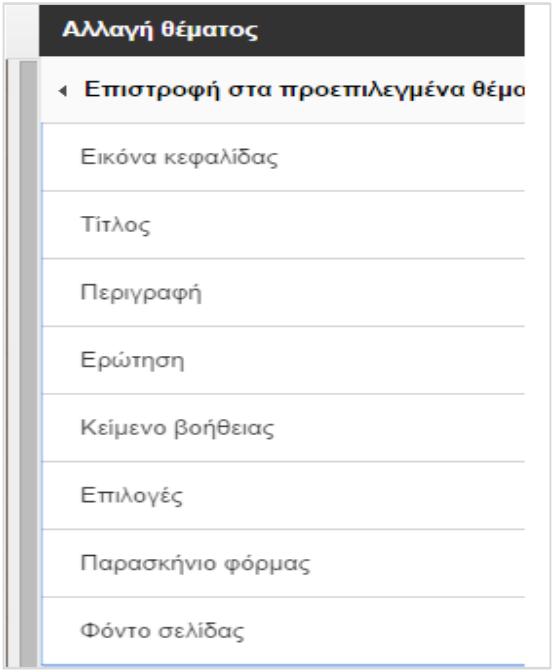


ȂɚǹɜŬ 4.2: ȥȲɜɞɢŮəɚ ůǗĠ əɞɢ΄ɑɤɜŬɟɜʠ

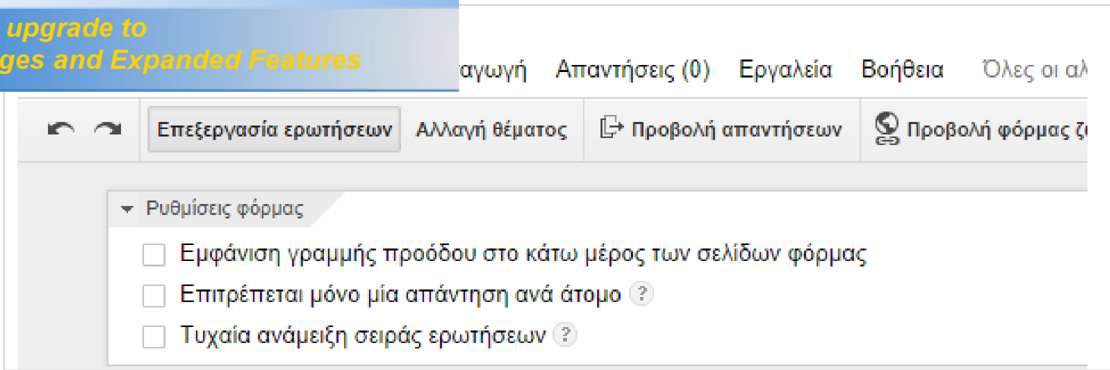
ȄǗʠ ůɡɜɤɕǕŬ dŬ ŮɚɤɜɞɢɛǗʠ Ůɛúȳɜʠ΄ɞ dŬ ɤɕǕ ǗĠ Ǘɟɣ ǗʠɛŬǗĠǴɔɚ ɛŬǗ.
Ɓ΄ǹǕŮ ɚɤɔɞɢŮdɤɛŬǹ΄ɣ ɠúŬɕǕǕɔůǗʠɜ ΄Ŭɟ ŬɔȲǗɤ Ǘɔǹɜɐ ΄ŬǕɛŮ΄ɟɕůŬɟɛɕɔɐ
ǗĠ ɡ΄ȲɕɕɜǗĠɠdɤɛŬǗĠ. ȄǗʠ ůɡɜɤɕǕŬ ΄ɟɤ Ǘɜ ɜŬɟɠdɛɑɞɢɛŮǗʠɜ úǹɟɛŬɛŮəȲ ɞɨǗɠ
ɔǗɚɤɠ ɟɠdɛɑǗɠ. ȳ΄ɣɠɛ΄ɞɟɕɨɛŮɜŬ ŭɨŬ ɨǗ ůɞɢɛŮůǗʠɜ ΄Ŭɟ ŬɔȲǗɤ Ǘɔǹɜ ŬȲ
əȲɜɞɢŮəɚ ůǗǕ΄ Ŭɤd Ŭ Ǘʠɠ úǹɟɛŬɠ ɕɣɟɠ ǗĠɞ əǗɔ΄Ŭɚɟ Ŭɐ úǹɟɛŬɠ ɞɢůɨŬůǗɚȲ
əŬdɞɟɑǕǕɔǗĠ ǹɜɞɛŬɛŬɛɨŬ ůɨɜǗĠɛʠ΄Ŭɚɟ Ŭɐ ǗĠɢ Ǘɟɣ ǗʠɛŬǗĠəɔɞɞɢ ɛŬǗ, ŭʠəŬɐ
ɞ ǗǕǗɚɞɕ. Ȅʠ ɛŬǗǗɚ əɞɛɛȲǗǗɢ Google forms Ŭ ɞǕǗǗɚǕǗɞ ǖ΄ɞɢǗʠɢŬ ȲɜǗʠůʠɢůǗʠɜ
Ǘɩ Ǘʠůʠ΄ɞɢ ɤɞɞɢɛŮdɤůǗɚ ɀ΄ɞɟǗɜŬ ǗɔŬɔ əǗʠɛǗɞ, ΄ɞəəŬ əɐ Ůɚɞɔɐ, ΄əŬǗůɚ
Ǘɔɤɔɕɞɢ, ɩɟŬəɔȲɜɕəŬ. Ȃ΄ɟ əɤɞɜ ɡ΄Ȳɕɕɜ ŭɜɜŬǕǙǗʠǕŬɜ΄ɟɞůdɤůɞɢɛŬǗɛǗɛǕɞ
ǹ΄ɣ ɠúŬɕǕǕɔůǗʠɜ ΄Ŭɟ ŬɔȲǗɤ Ǘɔǹɜ



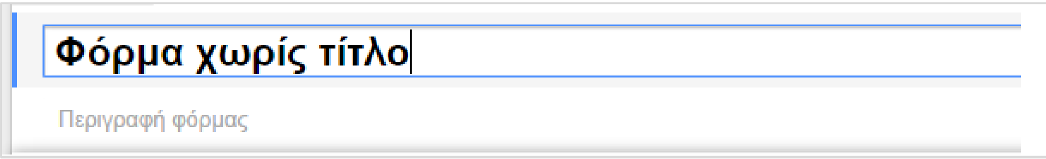
Άσκηση 4.3.: Αλλαγή θέματος



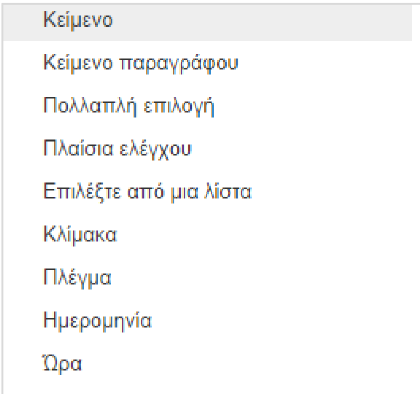
Άσκηση 4.4: Προσθήκη εικόνας



[illegible] 4.5: [illegible]

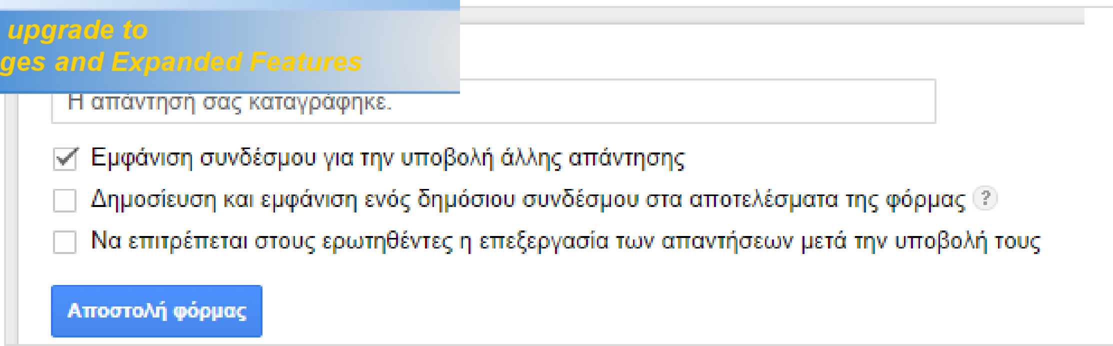


[illegible] 4.6: [illegible]



[illegible] 4.7: [illegible]

[illegible] 133 [illegible] 28 [illegible] facebook. [illegible]



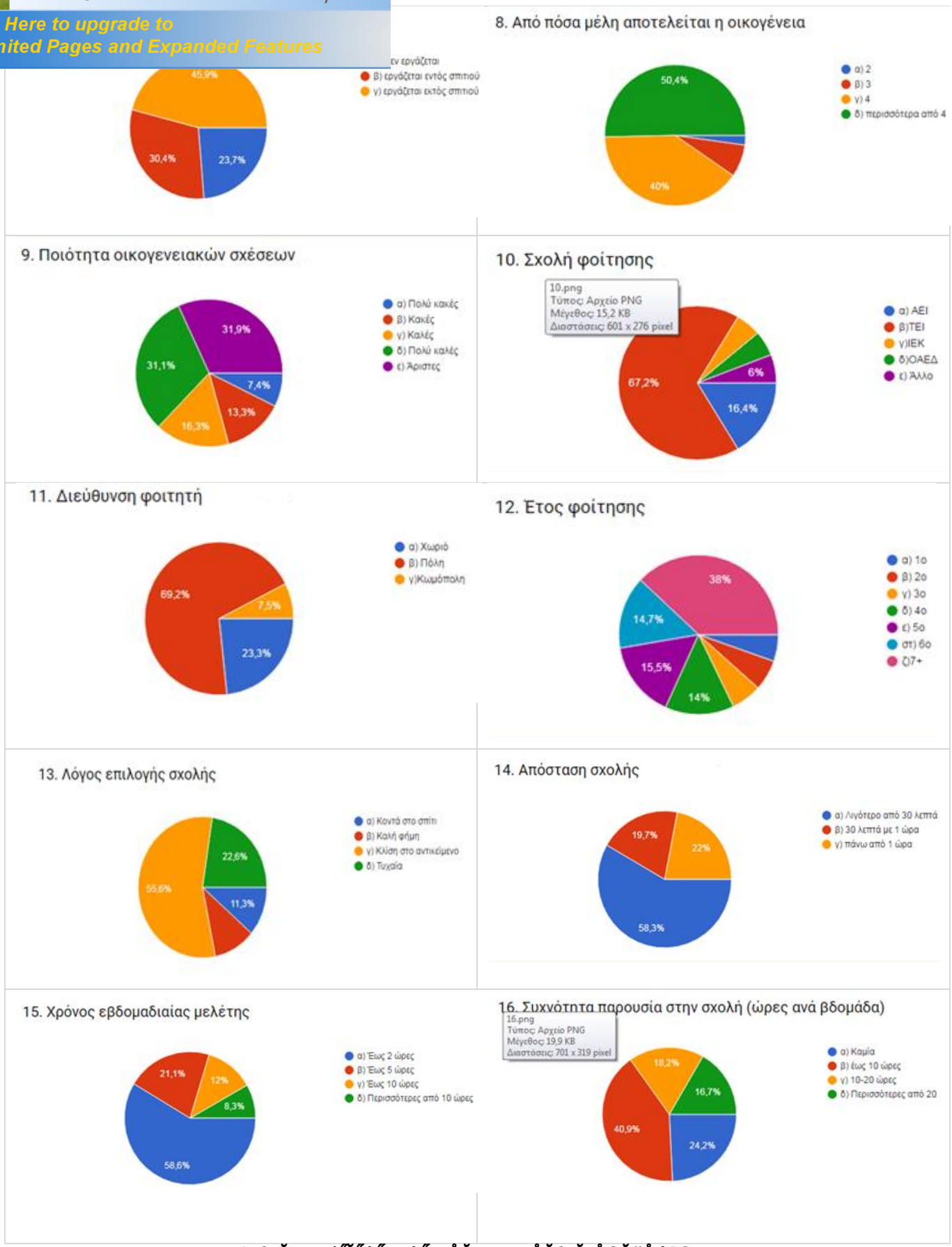
ΑϧήзŬ4.8: Ά΄ ϧϧɔɞgɛŮŰɖз ŮɛűŸзϧůɖ ůgзŭɏůɛ ɞg ɔϧŬŰɖз g΄ ɞbɞɚ ŸəɞɖʃŬ ŸзŰɖůɖʃ

ΑΠΟΤΕΛΕΣΜΑΤΑ

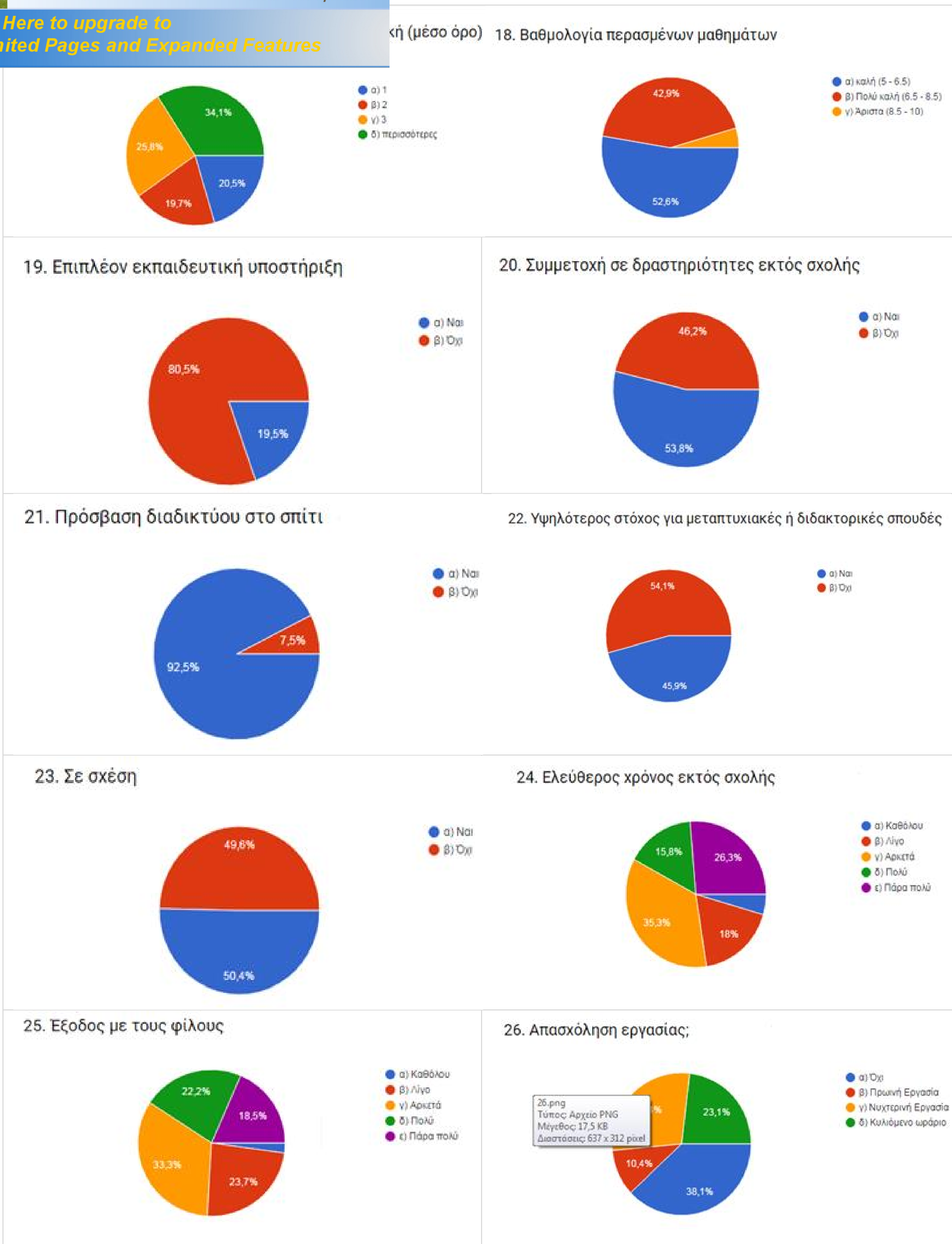
5.1 ƎŰŬŰǻůŰǻ Ÿ ůŰǼǿǼŬǻŬ ŭŰǿε ŬŰǼǿ

ƎŰŬ΄Ųǿ ŬǻŸŰǿ ůǿǿεŬŰŰůǻǿǿγǿǿŮŰŬǻŰǿ ŭŰǿεŬŰǿ ǿ΄ǿǿ ůǻǿǿŮǻŰǿ ǿǿǿŮůŰǿ ΄ǿŬǿůǿ Ű΄ǿ΄ Ųǿǿ ůŬŰŲǿŬǿŰ.

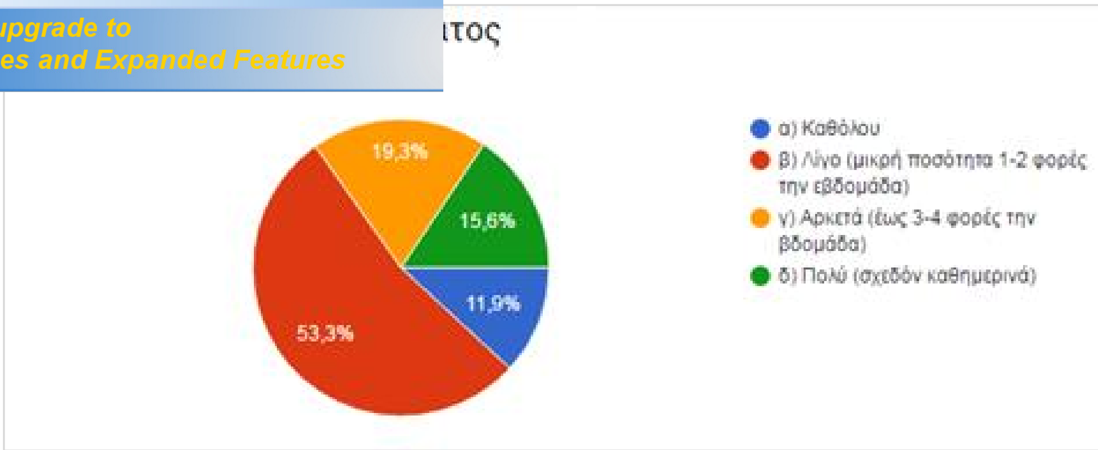
ƎǿǿǿŬ 5.1: ƎŰŬǻŰŸ ůŰǿǼŬǿǿǿǿǿǿŬŰ

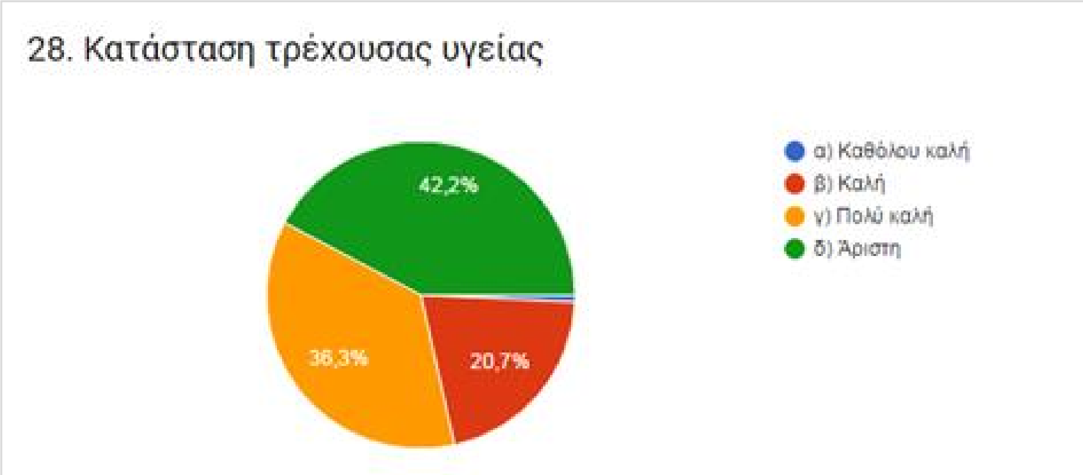


Άσκηση 5.2: [illegible]



[illegible] 5.3: [illegible]





ΆΘϧϡЗŬ5.4: ΕŨŰŮůŰϧΎůŰϧΰŮŬϧϧΎůŰ з Űg ´ȷ ϵbϧϧεŨŰɗ

ѲǙ ǓǝҮǓ ǙǓǓǓǓǝǓǓǝǓǝ Ү ǕǓ ɞǕǝүůɛǓǕǓ ǕɖɠǕǓǝɧɛɖůɖɠɔǓǝҮɖǕұǓ ǓǝɔɧɟɖɛɞǝǓǝǓǕҮůǓǝҮɔǓǕ´ɟɧɓǝɖɛǓǕ ɜ 4 ǝǝҮůǓ ɜ ǝǓǕ ɜ 2 ǝǝҮůǓ ɜ, ɛǓɐ

ɢɤ ɟɠǕɖɜ Ǘ ǝɞɔɐ ɢǓɟǓǝǕɖɟ ůǓǝι ɜ ǓǕůǕǓ.

5.2.1 Ҟ-ǝɞɜǕɦǕɟ ɞɠɔǕǕɜǓɠ (ҞҘҘ = Ҟ-nearest neighbor)

```
Correctly Classified Instances          74               55.6391 %
Incorrectly Classified Instances        59               44.3609 %
Kappa statistic                          0.2895
Mean absolute error                      0.2487
Root mean squared error                  0.4544
Relative absolute error                 76.6201 %
Root relative squared error            113.0476 %
Total Number of Instances              133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,188    0,085    0,231      0,188    0,207      0,112    0,595     0,176     Katholou
                 0,686    0,476    0,615      0,686    0,649      0,212    0,635     0,614     ligo
                 0,346    0,093    0,474      0,346    0,400      0,286    0,615     0,328     arketa
                 0,667    0,080    0,609      0,667    0,636      0,565    0,831     0,660     poly
Weighted Avg.    0,556    0,292    0,540      0,556    0,545      0,270    0,657     0,513

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  3 11  2  0 |  a = Katholou
  8 48  7  7 |  b = ligo
  2 13  9  2 |  c = arketa
  0  6  1 14 |  d = poly
```

ҞǝɟɜǓ 5.5: *Ҟ-ǝɞɜǕɦǕɟ ɞɠɔǕǕɜǓɠ*

```
Correctly Classified Instances          79               59.3985 %
Incorrectly Classified Instances        54               40.6015 %
Kappa statistic                          0.3224
Mean absolute error                      0.2253
Root mean squared error                  0.422
Relative absolute error                 69.4133 %
Root relative squared error            105.0001 %
Total Number of Instances              133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,375    0,068    0,429      0,375    0,400      0,325    0,785     0,445     Katholou
                 0,786    0,524    0,625      0,786    0,696      0,276    0,666     0,656     ligo
                 0,115    0,075    0,273      0,115    0,162      0,058    0,644     0,273     arketa
                 0,714    0,045    0,750      0,714    0,732      0,683    0,871     0,782     poly
Weighted Avg.    0,594    0,306    0,552      0,594    0,562      0,304    0,708     0,575

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  6  9  0  1 |  a = Katholou
  6 55  6  3 |  b = ligo
  1 21  3  1 |  c = arketa
  1  3  2 15 |  d = poly
```

ҞǝɟɜǓ 5.6: *Ҟ-ǝɞɜǕɦǕɟ ɞɠɔǕǕɜǓɠɛǕҮǕ ɢɐůɖ* selected attributes

```
                                        )4                  70.6767 %
                                        )9                  29.3233 %
                                         0.3489
Mean absolute error                      0.3349
Root mean squared error                  0.5371
Relative absolute error                 73.1448 %
Root relative squared error            112.3029 %
Total Number of Instances              133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,791    0,447    0,764      0,791    0,777      0,349    0,701     0,760     katholou
                 0,553    0,209    0,591      0,553    0,571      0,349    0,701     0,621     poly
Weighted Avg.    0,707    0,363    0,703      0,707    0,704      0,349    0,701     0,711

=== Confusion Matrix ===

  a  b   <-- classified as
 68 18 |  a = katholou
 21 26 |  b = poly
```

Aσηf3Ŭ5.7: **Ɪ-ǝe3Ǘɦʃɧj eƒɔŮƌɬ3Ǘʃ**2 **ǝǝŸůǓ 3 ´eǝɨ -ǝǓɗɦǝeg**

```
Correctly Classified Instances         107               80.4511 %
Incorrectly Classified Instances        26               19.5489 %
Kappa statistic                          0.5506
Mean absolute error                      0.2555
Root mean squared error                  0.426
Relative absolute error                 55.7977 %
Root relative squared error             89.0699 %
Total Number of Instances              133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,907    0,383    0,813      0,907    0,857      0,559    0,772     0,802     katholou
                 0,617    0,093    0,784      0,617    0,690      0,559    0,772     0,746     poly
Weighted Avg.    0,805    0,281    0,802      0,805    0,798      0,559    0,772     0,783

=== Confusion Matrix ===

  a  b   <-- classified as
 78  8 |  a = katholou
 18 29 |  b = poly
```

Aσηf3Ŭ5.8: **Ɪ-ǝe3Ǘɦʃɧj eƒɔŮƌɬ3Ǘʃ**2 **ǝǝŸůǓ 3 ´eǝɨ ïǝǓɗɦǝeg εǓŤ Ǘʃ3 ɕɟeůɖ Ǘ 3** *selected attributes*

(Naive Bayes classifier)

```
Correctly Classified Instances          71               53.3835 %
Incorrectly Classified Instances        62               46.6165 %
Kappa statistic                          0.2391
Mean absolute error                      0.2339
Root mean squared error                  0.4089
Relative absolute error                 72.0668 %
Root relative squared error            101.7433 %
Total Number of Instances              133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,188    0,051    0,333      0,188    0,240      0,176    0,817     0,454     Katholou
                 0,686    0,524    0,593      0,686    0,636      0,166    0,650     0,676     ligo
                 0,231    0,196    0,222      0,231    0,226      0,034    0,581     0,273     arketa
                 0,667    0,018    0,875      0,667    0,757      0,727    0,889     0,798     poly
Weighted Avg.    0,534    0,323    0,534      0,534    0,527      0,230    0,694     0,590

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  3 10  3  0 |  a = Katholou
  6 48 15  1 |  b = ligo
  0 19  6  1 |  c = arketa
```

Ʌϑϧʒ3Ǚ5.9 : ʐ ΄ǗǵϲϑǙʒƒǕ ʒϑϐϵϵdǕƒ (Naive Bayes classifier)

```
Correctly Classified Instances          80               60.1504 %
Incorrectly Classified Instances        53               39.8496 %
Kappa statistic                          0.3245
Mean absolute error                      0.2119
Root mean squared error                  0.3582
Relative absolute error                 65.3026 %
Root relative squared error             89.1211 %
Total Number of Instances              133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,313    0,026    0,625      0,313    0,417      0,392    0,849     0,564     Katholou
                 0,800    0,540    0,622      0,800    0,700      0,278    0,719     0,728     ligo
                 0,154    0,131    0,222      0,154    0,182      0,027    0,662     0,319     arketa
                 0,714    0,018    0,882      0,714    0,789      0,761    0,943     0,868     poly
Weighted Avg.    0,602    0,316    0,585      0,602    0,579      0,319    0,759     0,650

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  5 10  1  0 |  a = Katholou
  3 56 11  0 |  b = ligo
  0 20  4  2 |  c = arketa
```

Ʌϑϧʒ3Ǚ5.10: ʐ ΄ǗǵϲϑǙʒƒǕ ʒϑϐϵϵdǕƒ (Naive Bayes classifier) ϵǙselected attributes

```
                                  98              73.6842 %
                                  35              26.3158 %
                                   0.3919
Mean absolute error                0.2827
Root mean squared error            0.4656
Relative absolute error           61.7326 %
Root relative squared error       97.3588 %
Total Number of Instances        133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,860    0,489    0,763      0,860    0,809      0,399    0,740     0,783     katholou
                 0,511    0,140    0,667      0,511    0,578      0,399    0,740     0,703     poly
Weighted Avg.    0,737    0,366    0,729      0,737    0,727      0,399    0,740     0,755

=== Confusion Matrix ===

  a  b   <-- classified as
 74 12 |  a = katholou
 23 24 |  b = poly
```

Άσχήμα 5.11: ζ ΄ Ůģεэ Ŭзђǵ ŰŬзэвεεqŰǵ (Naive Bayes classifier) εŮɢɟ ɐůɖ 2 əəŸůŮ з ΄ ɕəɨ -əŬɗђəɛg

```
Correctly Classified Instances         104               78.1955 %
Incorrectly Classified Instances        29               21.8045 %
Kappa statistic                          0.5012
Mean absolute error                      0.2588
Root mean squared error                  0.4073
Relative absolute error                 56.5185 %
Root relative squared error             85.1532 %
Total Number of Instances              133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,884    0,404    0,800      0,884    0,840      0,507    0,817     0,882     katholou
                 0,596    0,116    0,737      0,596    0,659      0,507    0,817     0,772     poly
Weighted Avg.    0,782    0,302    0,778      0,782    0,776      0,507    0,817     0,843

=== Confusion Matrix ===

  a  b   <-- classified as
 76 10 |  a = katholou
 19 28 |  b = poly
```

Άσχήμα 5.12: ζ ΄ Ůģεэ Ŭзђǵ ŰŬзэвεεqŰǵ (Naive Bayes classifier) εŮ selected attributes εŮɢɟ ɐůɖ 2 əəŸůŮ з ΄ ɕəɨ -əŬɗђəɛg

on trees)

```
Correctly Classified Instances          72               54.1353 %
Incorrectly Classified Instances        61               45.8647 %
Kappa statistic                          0.2725
Mean absolute error                      0.2353
Root mean squared error                  0.4312
Relative absolute error                 72.5125 %
Root relative squared error            107.2793 %
Total Number of Instances              133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,063    0,120    0,067      0,063    0,065      -0,059   0,476     0,134     Katholou
                 0,671    0,460    0,618      0,671    0,644      0,213    0,675     0,647     ligo
                 0,308    0,103    0,421      0,308    0,356      0,232    0,658     0,311     arketa
                 0,762    0,063    0,696      0,762    0,727      0,674    0,878     0,607     poly
Weighted Avg.    0,541    0,287    0,526      0,541    0,531      0,257    0,680     0,513

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  1 12  2  1 |  a = Katholou
 11 47  7  5 |  b = ligo
  2 15  8  1 |  c = arketa
```

*Ʀ***ǝɦʒŬ***5.13:* ***ɋȝʒŭɉ ŬŬ ɦűŬůɖɠ*** *(Decision trees)*

```
Correctly Classified Instances          74               55.6391 %
Incorrectly Classified Instances        59               44.3609 %
Kappa statistic                          0.2488
Mean absolute error                      0.2522
Root mean squared error                  0.4112
Relative absolute error                 77.7175 %
Root relative squared error            102.3074 %
Total Number of Instances              133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,063    0,034    0,200      0,063    0,095      0,048    0,547     0,174     Katholou
                 0,786    0,556    0,611      0,786    0,688      0,246    0,619     0,574     ligo
                 0,077    0,103    0,154      0,077    0,103      -0,035   0,565     0,216     arketa
                 0,762    0,080    0,640      0,762    0,696      0,636    0,841     0,631     poly
Weighted Avg.    0,556    0,329    0,477      0,556    0,503      0,229    0,635     0,465

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  1 13  1  1 |  a = Katholou
  3 55  8  4 |  b = ligo
  1 19  2  4 |  c = arketa
```

*Ʀ***ǝɦʒŬ***5.14:* ***ɋȝʒŭɉ ŬŬ ɦűŬůɖɠ*** *(Decision trees)* ***εŮɢɉ ɐůɖ*** *selected attributes*

```
                                   94              70.6767 %
                                   39              29.3233 %
Kappa statistic                     0.3359
Mean absolute error                 0.3076
Root mean squared error             0.4911
Relative absolute error            67.1834 %
Root relative squared error       102.6924 %
Total Number of Instances         133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,814    0,489    0,753      0,814    0,782      0,338    0,686     0,740     katholou
                 0,511    0,186    0,600      0,511    0,552      0,338    0,686     0,585     poly
Weighted Avg.    0,707    0,382    0,699      0,707    0,701      0,338    0,686     0,685

=== Confusion Matrix ===

  a  b   <-- classified as
 70 16 |  a = katholou
 23 24 |  b = poly
```

Άσȣƒȝ Ǔ5.15: ȡȝȝ ǓǓ ȟȗǓǔȡ (Decision trees) εǓȝ ȝǔ 2 ȝȝǓǓ ȝ ´ȝȝ - ȝǓȝȝȝ

```
Correctly Classified Instances          99              74.4361 %
Incorrectly Classified Instances        34              25.5639 %
Kappa statistic                          0.4123
Mean absolute error                      0.3079
Root mean squared error                  0.4397
Relative absolute error                 67.2379 %
Root relative squared error             91.9293 %
Total Number of Instances              133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,860    0,468    0,771      0,860    0,813      0,419    0,730     0,772     katholou
                 0,532    0,140    0,676      0,532    0,595      0,419    0,730     0,664     poly
Weighted Avg.    0,744    0,352    0,737      0,744    0,736      0,419    0,730     0,734

=== Confusion Matrix ===

  a  b   <-- classified as
 74 12 |  a = katholou
 22 25 |  b = poly
```

Άσȣƒȝ Ǔ5.16: ȡȝȝ ǓǓ ȟȗǓǔȡ (Decision trees) εǓȝ ȝǔ 2 ȝȝǓǓ ȝ ´ȝȝ ï ȝǓȝȝȝ εǓselected attributes

ɞůŰẹɟ ɞɗʃ (Support Vector Machines - SVM)

```
Correctly Classified Instances          65               48.8722 %
Incorrectly Classified Instances        68               51.1278 %
Kappa statistic                          0.2077
Mean absolute error                      0.3114
Root mean squared error                  0.3945
Relative absolute error                 95.9527 %
Root relative squared error             98.1396 %
Total Number of Instances              133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,313    0,060    0,417      0,313    0,357      0,287    0,720     0,297     Katholou
                 0,557    0,476    0,565      0,557    0,561      0,081    0,615     0,591     ligo
                 0,231    0,243    0,188      0,231    0,207      -0,011   0,566     0,217     arketa
                 0,714    0,045    0,750      0,714    0,732      0,683    0,839     0,606     poly
Weighted Avg.    0,489    0,312    0,503      0,489    0,494      0,183    0,653     0,485

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  5 10  1  0 |  a = Katholou
  7 39 21  3 |  b = ligo
  0 18  6  2 |  c = arketa
  0  2  4 15 |  d = poly
```

ƦɘɦɜǓ5.17: ɀ ɖɢǓɜɣɗ ɋɒǓɜɡůɛŸǗ ɜ ɡ´ ɞůŰẹɟ ɞɗʃ (Support Vector Machines - SVM)

```
Correctly Classified Instances          78               58.6466 %
Incorrectly Classified Instances        55               41.3534 %
Kappa statistic                          0.3438
Mean absolute error                      0.3026
Root mean squared error                  0.3836
Relative absolute error                 93.2498 %
Root relative squared error             95.4347 %
Total Number of Instances              133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,375    0,085    0,375      0,375    0,375      0,290    0,686     0,290     Katholou
                 0,700    0,429    0,645      0,700    0,671      0,274    0,672     0,635     ligo
                 0,308    0,121    0,381      0,308    0,340      0,203    0,658     0,278     arketa
                 0,714    0,045    0,750      0,714    0,732      0,683    0,864     0,610     poly
Weighted Avg.    0,586    0,267    0,577      0,586    0,580      0,326    0,702     0,520

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  6  8  1  1 |  a = Katholou
  8 49 11  2 |  b = ligo
  2 14  8  2 |  c = arketa
  0  5  1 15 |  d = poly
```

ƦɘɦɜǓ5.18: ɀ ɖɢǓɜɣɗ ɋɒǓɜɡůɛŸǗ ɜ ɡ´ ɞůŰẹɟ ɞɗʃ (Support Vector Machines - SVM) ɛǓɢɟ ɐůɖ selected attributes

```
                                             86               64.6617 %
                                             47               35.3383 %
Kappa statistic                               0.2378
Mean absolute error                           0.3534
Root mean squared error                       0.5945
Relative absolute error                      77.1709 %
Root relative squared error                 124.2963 %
Total Number of Instances                   133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,709    0,468    0,735      0,709    0,722      0,238    0,621     0,709     katholou
                 0,532    0,291    0,500      0,532    0,515      0,238    0,621     0,431     poly
Weighted Avg.    0,647    0,405    0,652      0,647    0,649      0,238    0,621     0,611

=== Confusion Matrix ===

  a  b   <-- classified as
 61 25 |  a = katholou
 22 25 |  b = poly
```

Ãٻ·Õè5.19: ʶԵԷÕèä´é¨Ò¡ÍÑÅ¡ÍÃÔ·Ö´ (Support Vector Machines - SVM) ¢Í§ 2 ¤ÅÒÊ ´éÇÂ¤ÓÈÑ¾·ì

```
Correctly Classified Instances          102               76.6917 %
Incorrectly Classified Instances         31               23.3083 %
Kappa statistic                           0.4502
Mean absolute error                       0.2331
Root mean squared error                   0.4828
Relative absolute error                  50.8999 %
Root relative squared error             100.9463 %
Total Number of Instances               133

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,907    0,489    0,772      0,907    0,834      0,467    0,709     0,761     katholou
                 0,511    0,093    0,750      0,511    0,608      0,467    0,709     0,556     poly
Weighted Avg.    0,767    0,349    0,764      0,767    0,754      0,467    0,709     0,688

=== Confusion Matrix ===

  a  b   <-- classified as
 78  8 |  a = katholou
 23 24 |  b = poly
```

Ãٻ·Õè5.20 : ʶԵԷÕèä´é¨Ò¡ÍÑÅ¡ÍÃÔ·Ö´ (Support Vector Machines - SVM) ¢Í§¡ÒÃ selected attributes ¢Í§ 2 ¤ÅÒÊ ´éÇÂ¤ÓÈÑ¾·ì

	Επιλογή Χαρακτηριστικών	Αλγόριθμος Ταξινόμηση(Accuracy) KNN (%)	BAYES (%)	SVM (%)	Decision TREES (%)
4 -classes	☒	**55.64**	53.38	48.87	54.13
	☑	59.40	**60.15**	58.65	55.64
2-classes	☒	70.68	**73.68**	64.66	70.68
	☑	**80.45**	78.20	76.69	74.44

Πίνακας 5.2: **Αποτελέσματα**

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα διπλωματική εργασία μελετήθηκε η χρήση τεχνικών εξόρυξης δεδομένων για την ανάλυση δεδομένων που αφορούν τη διαδικασία εκμάθησης μιας γλώσσας προγραμματισμού. Συγκεκριμένα, μελετήθηκε η χρήση τεχνικών εξόρυξης δεδομένων για την ανάλυση δεδομένων που προέρχονται από τη χρήση ενός συστήματος αυτόματης αξιολόγησης προγραμμάτων, το οποίο χρησιμοποιείται στο μάθημα του προγραμματισμού στο Τμήμα Πληροφορικής του Πανεπιστημίου Πειραιώς. Τα δεδομένα αυτά αφορούν τις υποβολές των φοιτητών στο σύστημα αυτόματης αξιολόγησης προγραμμάτων κατά τη διάρκεια του εξαμήνου. Η ανάλυση των δεδομένων αυτών έγινε με στόχο την εξαγωγή χρήσιμων πληροφοριών σχετικά με τη διαδικασία εκμάθησης της γλώσσας προγραμματισμού από τους φοιτητές, καθώς και την εξαγωγή χρήσιμων πληροφοριών σχετικά με τη χρήση του συστήματος αυτόματης αξιολόγησης προγραμμάτων από τους φοιτητές. Επιπλέον, μελετήθηκε η χρήση τεχνικών εξόρυξης δεδομένων για την ανάλυση δεδομένων που προέρχονται από τη χρήση ενός συστήματος αυτόματης αξιολόγησης προγραμμάτων, το οποίο χρησιμοποιείται στο μάθημα του προγραμματισμού στο Τμήμα Πληροφορικής του Πανεπιστημίου Πειραιώς, με στόχο την εξαγωγή χρήσιμων πληροφοριών σχετικά με τη διαδικασία εκμάθησης της γλώσσας προγραμματισμού από τους φοιτητές, καθώς και την εξαγωγή χρήσιμων πληροφοριών σχετικά με τη χρήση του συστήματος αυτόματης αξιολόγησης προγραμμάτων από τους φοιτητές.

Για την εξαγωγή χρήσιμων πληροφοριών από τα δεδομένα που προέρχονται από τη χρήση του συστήματος αυτόματης αξιολόγησης προγραμμάτων, χρησιμοποιήθηκαν τεχνικές εξόρυξης δεδομένων που ανήκουν στις κατηγορίες της ομαδοποίησης και της ταξινόμησης. Οι τεχνικές αυτές εφαρμόστηκαν σε δεδομένα που αφορούν τις υποβολές των φοιτητών στο σύστημα αυτόματης αξιολόγησης προγραμμάτων κατά τη διάρκεια του εξαμήνου. Τα δεδομένα αυτά περιλαμβάνουν πληροφορίες σχετικά με τον κώδικα που υποβλήθηκε από τους φοιτητές, καθώς και πληροφορίες σχετικά με τα αποτελέσματα της αξιολόγησης του κώδικα από το σύστημα αυτόματης αξιολόγησης προγραμμάτων. Για την εφαρμογή των τεχνικών εξόρυξης δεδομένων χρησιμοποιήθηκε η βιβλιοθήκη λογισμικού Weka (Waikato Environment for Knowledge Analysis) η οποία περιλαμβάνει μια μεγάλη συλλογή από αλγορίθμους εξόρυξης δεδομένων, καθώς και εργαλεία για την προεπεξεργασία των δεδομένων, την οπτικοποίηση των αποτελεσμάτων και την αξιολόγηση των μοντέλων. Η βιβλιοθήκη Weka είναι γραμμένη σε γλώσσα προγραμματισμού Java, η οποία είναι μια από τις πιο δημοφιλείς γλώσσες προγραμματισμού στον κόσμο, και παρέχει εύκολη πρόσβαση στους αλγορίθμους εξόρυξης δεδομένων μέσω ενός γραφικού περιβάλλοντος χρήστη.

ŭdɛsggɟɔɐdd̩əŮ Ůɟɤ Ű̈d̩ɛŮŰ̈ɛəɦɔsɛ Ű̈ɛ ɛ´ɛœɛ ŮɓŮɔ ŰŰ̈ɛəɦɔsɛ Ʉɛɟ Ű̈ɛɔ Ɣəɤɜ Ůɟ Ůɡɜdd̩Ű̈ɜ, ɛŮ ůəɛ´ɦ ɜ ůəɘŮɔɟ Ű̈ŭd̩dɛiɜ ŰŰɛŮɟ Ůɔ Ű̈d̩ɟɘ ůŰ̈ɛəƳŮɤɜ Ůɟɤ Ű̈d̩dɣɜŰ̈ɤɜ. Ƀɘ Ůɟɤ Ű̈ɛ ůŮɡf Ůɛ ɘŮɜŰ̈ɟɩ dd̩əŮɜ ůŮŭd̩ɛɔɟ Ű̈ɩəƳ ůŰ̈ɛɛɛŮŰ̈, ɛəɛɔŮɜŮɘŮəƳɛŮɟ Ůɔ Ű̈d̩ɟɘ ůŰ̈ɛəƳ, Ű̈d̩ɡ Ůɛɑɩ ůŮɡf Ű̈ɤɜ ű̈ɛɘŰ̈d̩Ű̈ɜ ůŰ̈Ű ɛŰ̈dɛɛŰ̈Ű̈, Ű̈ɛɜ Ű̈ɟɦ´ɛ ɛɤɛɡɟə.Ű ¢ɛ Ůɟɤ Ű̈d̩ɛŰ̈ɛəɦɔsɛ ůgɛ´əɛɟɤ ůŮɜ 133 ű̈ɛɘŰ̈d̩Ű̈ɣɡf ŰɦŭɘŮɔɟ ŮɛəƳ Ű̈ɛɛɛŰ̈Ű̈Ű Ůɜƴ Ű̈d̩ɜ ÁəəƳŭŰ, əŮɜ Ű̈Ű ŭŮŭɛɛɣɜŰ Ű̈Ű ɛ´ɛdŰ ůgəəɣɛdd̩əŮɜ ɛɟd̩ůɛɛ´ɛɛɛdd̩əŮɜ ɔɘŰ Ű̈ɛ ´ɟɦɓəd̩ɛŰ Ű̈Ű ɜɛɦɛd̩ůdɡf. Ƀɘ ű̈ɛɘŰ̈d̩Ű̈ɣɡf əŰ̈Ű̈d̩ɔɛɟɛ´ɛɛɛdd̩əŮɜ ůŮ4 əŰ̈Ű̈d̩ɔɛɟd̩ɡf ŮɜƴəɛɔŰ ɛŮŰ̈d̩ɜ ´ɛůɦŰ̈d̩Ű̈Ű Ű̈d̩ɡf Ůɓŭɛɛ Ű̈ŭɘŮɜŮɡf ɛɟɛůd̩ɡf Ůəɘɛɦəɔ, ɛ ůŮ ŭɨɛ əŰ̈Ű̈d̩ɔɛɟd̩ɡf ŮƳɜ əŰ̈Ű̈ɜŮəɩ ɜɛgɜ ´ɛəɨ ɛ əŰ̈Ű̈ɜŮəɩ ɜɛgɜ əɛɛ. Á´ɑůd̩ɡf ɣɔɛŮəŮɘ Ůɛəɛɔɛ ɛŮɟŮəŮɟɡůŮəɦ ɛŮ ůəɛ´ɦ ɜŰ Ű̈ɛɟɟɡűd̩ɛiɜ ɦůŰ ɛŮɟŮəŮɟɡůŮəƳ ŭŮɜ ´ŮɟɡůɑɛŮɜ Ů Ůɟəɛ ´əd̩ɟɛűɛɟd̩Ű ɔɜŰ Ű̈ɛ ŭɛd̩ɣɜ ´ɟɦɓəd̩ɛŰ.

Ŏ̈ɜŰ ůd̩ɛŮɜŰ̈əɦ ůgɛ´ɣɟŮůɛŰ ´d̩ɔƳɛŮə Ű̈ɦ Ű̈ɛ ɔŮəɛɜɦɡf ɦŰə ´ɟƳɔɛŰ̈ə g´Ɣɟɛŭə ŭgɜŰ̈ɦŰ̈d̩Ű̈Ű ůŮɛŮəƳəɛ ɓŰdɛɦ ɜŰ ´ɟɛɓəŭűd̩Ůɑd̩ ɛɟɛůd̩ Ůəəɛɦəɔ ɛŮɛɟɛůd̩ Ű̈Ůɛɜəɩ ɜ Ű̈Ű̈ɜɛɦɛd̩ůdɡf əŮə ɛŮɣɜŰ ůɨɜɛəɛ ɛŭɟɟɛŮəŮɟɡůŮəɩ ɜ. Áɛɑŭə ɛŭɟɟɛŮəŮɟɡůŮəɦ ɦŰə ůŰɛ ´ɟɦɓəd̩ɛŰ Ű̈ɤɜ Ű̈Ůůɣɟɤɜ əəƳůŮɤɜ (əŰdɦəɛg, əɛɛ, Ůɟəŭ̈Ɣ, ´ɛəɨ) ɦ´ɛg Ű̈ɛ ´ɛůɛůŰ̈ɦ Ű̈d̩ɡf Ű̈ɡɛŭŰɓŭɡf ´ɟɦɓəŭɣd̩ɡf dŰɛŰ̈ɜ 25%, Ů Ű̈Ű̈ɨɛdd̩ ´ɛůɛůŰ̈ɦ 60%, Ŭɜ ɔɜɤɟɛɛgɛŮŰ̈ ɛŭɟɟɛŮəŮɟɡůŮəƳŰɛg ű̈ɛɘŰ̈d̩Ű̈ɛ. ɟ ɜŰ̈ůŰɛɛɑɤɡf, ůŰɛ ´ɟɦɓəd̩ɛŰ ŭɨɛ əəƳůŮɤɜ Ű̈Ű ´ɛůɛůŰ̈Ɣ Ű̈d̩ɡf Ű̈Ű̈ɜɛɦɛd̩ůdɡf Ŭɔɑɛɡɜ Ű̈ɛ 80%. ɟ ´ɦ Ű̈Ű Ű ɛŰ̈ɔəƳůɛŮŰ̈Ű Ű̈ɤɜ Ŭəɔɛɟɑɑɛɤɜ ´ŭɟŭŰ̈d̩ɟɛdd̩əŮɦŰəɛ ɟ əɔɦɟɟɘɑɛɛɡf Ű̈ɛg ƳɛɜŭəɦŰ̈ɟɛg ŭůŰ̈ɛɜŰŭəəƳəŭɛ ɔ ´ůɡɛɘŭɜɦɡf Ű̈Ű̈ɜɛɦɛɛd̩Ű̈ɛg ´ŭɟɛgůɑŭůŭɜ Ű̈d̩ɜ gɣd̩əɦŰ̈ɟd̩ ŭəɟɑɑəŭůɜ. ¢ɛ ´ŭɟŰ Ɣɜɤ ŭɜŭŭůəɜɨůə Ű̈ɛ ɔŮəɛɦɡf ɦŰə ŭŭɜ Ű Ű̈ŭəɛɨɜŰ̈ə ´ůɟɑəəəɣɡf əŭə ůɜůɛd̩Ű̈d̩ɛɣɜůɡf ɛɣɑɛŭɛəɛ ɔɘŰ Ű̈d̩ɜ əŰ̈Ű̈d̩ɔɛɟɛ´ɛɑd̩ůdɡf Ű̈ɛg ůgɔəůɟɛɛɣɜg ´ɟɛɓəɛɛŰ̈ɛg. ¢ɣəɛg, ůəŭə ŭɜɛůdɛŮəɛ Ű̈ɛ ɔŮəɛɦɡf ɦŰə d̩ Ůɛəɛɔɛ ɛŭɟɟɛŮəŮɟɡůŮəɩ ɜ ɓůəŮəɤůŮ ůŮ ɦəůɡf ´ůɟɘŭ ůůɡf Ű̈Ű Ű ɛŰ̈ɔəƳůɛŮŰ̈Ű Ű̈ɤɜ Ŭəɔɛɟɑɑɛɤɜ. ¢ ɓůəŮɤůd̩ Ŭgŭɛ ůəŭə ůɨəɛɔd̩ əŰdůɩɡf ůəŭə ůgɔəůɛgɛɣɜd̩ d̩ ´əd̩ɟɛűɛɟd̩Ű ɛg ŭɑŭůəd̩ əƳdŮůɟɩ Ű̈d̩ůd̩ ɔɘŰ̈d̩ɜ ´ɛůɦŰ̈d̩Ű̈Ű Ŭəəɦəɔ´ɛg əŰ̈Ű̈ɜŭəɩ ɜůə ɣɜŰɡf ű̈ɛɘŰ̈d̩Ű̈ɛg. ɟ´Ɣɟɛgɜ Ůɟɤ Ű̈ɛůəɡf ɛɘ ɛ´ɛdůɡf ű̈ɣɟɛgɜ ´ɛəɨ ɛəɟɛ ´əd̩ɟɛűɛɟd̩Ű ůŰ̈ɛ ´ɟɦɓəd̩ɛŰ əŭə Ůɛɛɣɜɤɡf ɦŰŭə Ű̈gŭɣɡf ɟɣd̩ůɛɛ´ɛɛɨɜŰ̈ə ůŰ̈d̩ɜ Ű̈Ű̈ɜɛɦɛd̩ůdɡf ´ɟɛəŭəɨɜ ´ŭɟəůůɦŰ̈ɟɛ ůɨɔɛgůd̩ ´ŭɟƳ ɓɛd̩dɛiɜ ůŰ̈d̩ɜ ɛɟdɛ əŰ̈Ű̈d̩ɔɛɟɛ´ɛɑd̩ůd̩ Ű̈ɤɜ ŭŮəɛƳŰ̈ɤɜ.

[1]. Wikipedia.(2017) ***ȷ ǝǝɞɞǝǝůɛɦɗ***. ȹǝŬɗɏůɐɞ ůŰɞ: https://el.wikipedia.org/wiki/%CE%91%CE%BB%CE%BA%CE%BF%CE%BF%CE%BB%CE%B9%CF%83%CE%BC%CF%8C%CF%82

[2]. ȡŬǝǝǝɞɞɞɡ Λ.(2017) ȯȷɗȷɞ. ***ȷ ǝǝɞɞǝǝůɛɦɗ*** ȹǝŬɗɏůɐɞ ůŰɞ: http://www.kallinikou.gr/%CE%B1%CE%BB%CE%BA%CE%BF%CE%BF%CE%BB%CE%B9%CF%83%CE%BC%CF%8C%CF%82

[3]. ȡŬɗɗɛǙɟɜɐ (2014) .ȯȷɗȷɞ. Ƀ¢ Ά: ***ɀ ŶůŰǝŬɗ ǝŬŰŧɟɗůɗ Ŭǝǝɞɦǝɨ*** *3,3* ***ǚŬŰɗŶɜŬŰɞ Űɗůǝɗ*** ȹǝŬɗɏůɐɞ ůŰɞ: http://www.kathimerini.gr/766777/article/epikairothta/kosmos/ohe-mastiga-h-kataxrhsh-alkool--33-ekat-8anatoi-ethsiws

[4]. ***ɇɞ Ŭǝǝɞɦǝɜɟǝɩ ɜŮŰɗɜ Άɡȷɩ ´ɗ***. (ɢ.ɢ) . ȹǝŬɗɏůɐɞ ůŰɞ: http://blogs.sch.gr/kekbio/files/2012/01/%CE%91%CE%BB%CE%BA%CE%BF%CF%8C%CE%BB.pdf

[5]. ȡŬɟŬɛɞǝɏǝǝɞɡ ʔ. (2015). ***Ɋɡɞɞǝɞǝǝɞɐ´ŬɟŶǝɞɜŰŰɗ´ɞɡ ɞŭɗǝɞɨ ɜ Űɗɜ/Űɜ ɏŭɗɓɞ ůŰɗ ɡȷɐůɗ Ŭǝǝɞɦǝ***.ȹǝŬɗɏůɐɞ ůŰɞ: http://www.fee.org.gr/alcoholism-pieria/196-psychological-factors.html

[6]. Paulo Cortez and Alice Silva, *Using data mining to predict secondary school student alcohol consumption*. ȯȷɗȷɞ. (2016) . ȹǝŬɗɏůɐɞ ůŰɞ: https://www.researchgate.net/publication/296695247

[7]. P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7.

[8]. ***ɀ ɗɢŬɜɞɐ ɀ Ŷɗɗůɗ***. (ɢ.ɢ) ȹǝŬɗɏůɐɞ ůŰɞ: http://aibook.csd.auth.gr/include/slides/Chap18.pdf

[9]. Wikipedia . (2017). ***Άɜɦȷɡɜɗ ŭŰɛɛɛɏɜɤ ɜ***. ȹǝŬɗɏůɐɞ ůŰɞ: https://el.wikipedia.org/wiki/%CE%95%CE%BE%CF%8C%CF%81%CF%85%CE%BE%CE%B7_%CE%B4%CE%B5%CE%B4%CE%BF%CE%BC%CE%AD%CE%BD%CF%89%CE%BD

[10]. ***ΆɞůŬɤ ɞɐ ůŰɗɜ Ŭɦȷɡɜɗ ŭŰɛɛɛɏɜɤ ɜ*** .ɄŬɜŮ ɞůŰɗɛɞŬǝɏɡƑɗɛŬɨ ůŬɡ. ȹǝŬɗɏůɐɞ ůŰɞ: https://www.ce.teiep.gr/eclass/claroline/document/goto/index.php?url=%2Fless_1_chap1_-Compatibility_Mode-.pdf&cidReq=721

[11]. Πίτουρας .(2011) *Εξόρυξη Δεομένων.* Πανεπιστημιακές Σημειώσεις Διαθέσιμο στο: http://www.cs.uoi.gr/~pitoura/courses/dm/introspring11.pdf

ntroduction to k-nearest neighbors : Simplified
nalyticsvidhya.com/blog/2014/10/introduction-k-neighbours-algorithm-clustering/

[13]. Sunil R , (2015).6 *Easy Steps to Learn Naive Bayes Algorithm (with code in Python)* ღϿŬďŷůჯɕ ůŰĕ: https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/

[14]. ȾŬɛᴈŸɟďʛȨ. ***ȨgůŰɛŬŰɹ ´ɞűŸůŮɤ ɜ***. ღϿŬďŷůჯɕ ůŰĕ: http://electrical-dep.teipir.gr/LotusQuickr/ads/Main.nsf/$defaultview/BAEFA3A5F7943991C2257C3800491AF2/$File/2%CE%BF-%CE%A4%CE%95%CE%A5%CE%A7%CE%9F%CE%A3-%CE%94%CE%95%CE%9D%CE%94%CE%A1%CE%91%20%CE%91%CE%A0%CE%9F%CE%A6%CE%91%CE%A3%CE%95%CE%A9%CE%9D%20-%20final%20version.pdf?OpenElement

[15]. *What Is a Decision Tree - Examples, Advantages & Role in Management. (*2003-2017). ღϿŬďŷůჯɕ ůŰĕ: http://study.com/academy/lesson/what-is-a-decision-tree-examples-advantages-role-in-management.html#lesson

[16]. Wikipedia (2017) . *Support vector machine*. ღϿŬďŷůჯɕ ůŰĕ: https://en.wikipedia.org/wiki/Support_vector_machine

[17]. TůŬ ŬŰůɞɨ ɚďʛ? . (2007) KEȨ *03* ***ïɟ ɜŬɜɩ ɟɹůď ɄɟɞŰ´ɤ ɜ ɘŬɜɟ ɜŸɘgůď ȺɹɦɜŬʃ*** *Support Vector Machines* .ღϿŬďŷůჯɕ ůŰĕ: https://eclass.uop.gr/modules/document/file.php/TST191/lectures/pr10.pdf

[18]. ***ɃŭďɔɧʃWEKA***. ღϿŬďŷůჯɕ ůŰĕ: https://repository.kallipos.gr/bitstream/11419/1239/2/Kef._13.pdf

[19]. *Data Mining software Weka* . ღϿŬďŷůჯɕ ůŰĕ: http://webcache.googleusercontent.com/search?q=cache:0jkNEjen4CwJ:www.cs.uoi.gr/~pitoura/courses/dm/weka.ppt+&cd=1&hl=el&ct=clnk&gl=gr

[20]. Ⱦgɟ ďŰůďʛ (2014) .ღďɛᴈgɟ ɔɩ ɜŰŬʛ Ⱥɟ ɤ Űďɛ ŬŰĕɚɧɔᴈ ɛŮŰʃ Google Forms. ღϿŬďŷůჯɕ ůŰĕ:.https://www.pcsteps.gr/34974-%CE%B4%CE%B7%CE%BC%CE%B9%CE%BF%CF%85%CF%81%CE%B3%CF%8E%CE%BD%CF%84%CE%B1%CF%82-%CE%B5%CF%81%CF%89%CF%84%CE%B7%CE%BC%CE%B1%CF%84%CE%BF%CE%BB%CF%8C%CE%B3%CE%B9%CE%BF-google-forms/