Abstracte

Author Guidelines for 8.5 x 11-inch Proceedings Manuscripts

Pere Girona Campi

https://github.com/perot84

pere.girona@onmail.com

El projecte intenta fer una predicció de quan triga a trobar feina un usuari del servei d’inserció laboral d’una entitat que treballa amb persones. S’ha treballat amb diferents dades de bases extretes del CRM de la pròpia entitat i el s’ha obtingut un resultat molt limitat, degut en part a la poca quantitat i qualitat de les dades. Cal tenir en compte que la recollida de totes les dades és manual, la qual cosa pot provocar diferents criteris a l’hora d’introduir-les (tant com es recullen, com si es recullen o no).

# Introducció

Per a crear el conjunt de dades s’han extret diferents bases de dades a partir del CRM de l’entitat. Ens hem trobat amb un conjunt de dades força limitat en nombre i varis problemes en la consistència de dades, en la introducció d’aquestes i amb un nombre força elevat de valors faltants. Hem aprofitat aquesta situació per a ampliar els objectius prèviament descrits. Hem decidit crear una base de dades on unir totes les dades pertanyent a cada usuari, així com introduir i seleccionar característiques no recollides anteriorment, fer un recull de propostes de millora la de recollida i gestió de les dades, així com argumentar la necessitat d’una bona recollida de dades a partir de la utilitat que pot donar a les persones que les introdueixen. En aquest últim apartat cal destacar que qui fa més ús d’aquestes dades no són les persones tècniques que les introdueixen, si no que moltes vegades és personal tècnic superior, per aquest motiu hem pensat en aquest objectiu.

Entitat social d’on extraiem les dades i amb la qui treballem:

L’Associació Catalana per a la Promoció de les Persones amb Sordesa (ACAPPS), creada l’any 1992, és una entitat sense ànim de lucre declarada d’Utilitat Pública que té com a finalitat representar i defensar els interessos globals de les persones amb discapacitat auditiva i de les seves famílies. ACAPPS té com a àmbit territorial Catalunya. Les tres associacions (ACAPPS Barcelona, ACAPPS Lleida i ACAPPS Vallès) formen part de la Federació d’Associacions Catalanes de Pares i Persones amb Sordesa (ACAPPS).

La missió d’ACAPPS és representar i defensar els drets i interessos globals de les persones amb sordesa i de les seves famílies, a tots els nivells, davant la societat, administracions públiques i altres institucions, integrant i impulsant amb aquesta finalitat l’acció de les famílies, o representants legals, i de les persones amb sordesa.

# Metodologia

En aquest projecte el preprocessat ha tingut un paper molt important, ja que hem hagut d’utilitzar 7 bases de dades diferents amb múltiples errors, migracions incompletes, valors faltants, duplicats d’ids...

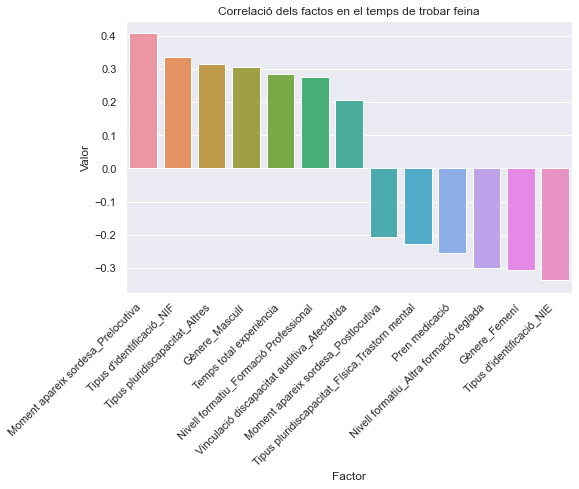
L’objectiu del preprocessat ha estat mantenir el màxim nombre de característiques per a millorar les mètriques alhora que s’han creat noves columnes. Destacar que s’ha creat la la columna de temps que triga la persona a trobar feina, la qual ha estat l’objectiu a preveure.

Hem provat diferents models regressius per a cercar el temps que triga un usuari a trobar feina en què el millor resultat ens l’ha donat KNeighborsRegressor.

Hem usat les mètriques R2, MSE i MAE. Degut a que han sortit uns resultats molt baixos ens hem guiat per la MAE, ja que també ens permetia entendre i transmetre millor el significat al client.

# Resultats

El millor resultat l’hem obtingut amb KNeighborsRegressor amb un resultat en el MAE de 87.6 dies. Per altra banda hem recollit els valors de correlació entre les característiques i la variable amb els següents resultats:



.

# Conclusions

Tot el procés del projecte ha estat centrat en oferir i donar utilitat a l’aprenentatge adquirit a una entitat social, sense contacte previ en la ciència de dades.

Encara que el resultat de la predicció ha estat limitat s’han pogut aportar diferents coneixements a l’entitat:

Millores de recollida de les dades

Millores en la gestió de les dades (CRM)

Coneixement i utilitat de la ciència de dades per a extraure més profit a les dades recollides.

Donar més valor a la introducció de les dades, la qual cosa pot repercutir en una millora de la quantitat i qualitat d’aquestes.

Per altra banda hem comprovat com és d’important explicar prèviament i de la millor manera les possibilitats del Machine Learning, ja que a mesura que han avançat les reunions hem pogut recollir més demandes del client i donar més possibilitats a la feina realitzada.

Per a que la utilitat fos més gran caldria crear una API per tal d’automatitzar la recollida de dades i la creació de les base de dades.

# Second and following pages

The second and following pages should begin 1.0 inch (2.54 cm) from the top edge. On all pages, the bottom margin should be 1-1/8 inches (2.86 cm) from the bottom edge of the page for 8.5 x 11-inch paper; for A4 paper, approximately 1-5/8 inches (4.13 cm) from the bottom edge of the page.

# Type-style and fonts

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 fonts are preferred.

# Main text

Type your main text in 10-point Times, single-spaced. Do **not** use double-spacing. All paragraphs should be indented 1/4 inch (approximately 0.5 cm). Be sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

**Figure and table captions** should be 10-point boldface Helvetica (or a similar sans-serif font). Callouts should be 9-point non-boldface Helvetica. Initially capitalize only the first word of each figure caption and table title. Figures and tables must be numbered separately. For example: “Figure 1. Database contexts”, “Table 1. Input data”. Figure captions are to be centered *below* the figures. Table titles are to be centered *above* the tables.

# First-order headings

For example, “1. Introduction”, should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after. Use a period (“.”) after the heading number, not a colon.

## Second-order headings

As in this heading, they should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after.

**8.1.1.Third-order headings.** Third-order headings, as in this paragraph, are discouraged. However, if you must use them, use 10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

# Footnotes

Use footnotes sparingly (or not at all) and place them at the bottom of the column on the page on which they are referenced. Use Times 8-point type, single-spaced. To help your readers, avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example 1. Where appropriate, include the name(s) of editors of referenced books 2.

# References

[1] A.B. Smith, C.D. Jones, and E.F. Roberts, Article Title, Publisher, Location, Date.

[2] Jones, C.D., A.B. Smith, and E.F. Roberts, Book Title, Publisher, Location, Date.

# Copyright forms and reprint orders

You must include your fully-completed, signed IEEE copyright release form when you submit your paper. We **must** have this form before your paper can be published in the proceedings. The copyright form is available as a Word file, <copyright.doc>, as a PDF version, <copyright.pdf>, and as a text file in <authguid.txt>.

Reprints may be ordered using the form provided as <reprint.doc> or <reprint.pdf>.