

Hoy te convertís en (Junior) Data Scientist

Ricardo Colombo Diego Santos Luis Badel Erik Machicado

*Departamento de Computación
Universidad de Buenos Aires
C.A.B.A, Argentina*

Abstract

El trabajo consiste en aplicar técnicas de Métodos numéricos y Data Science, en particular Regresiones Lineales y Cuadrados Mínimos, a un gran conjunto de datos reales, buscando identificar modelos capaces de predecir comportamientos .

Keywords: Data Science, KPIs, Cuadrados Minimos Lineales (CML)

1 Introduction

La industria aeronáutica , por su gran estructura organizativa, infraestructura, medidas de seguridad y costos operativos necesita llevar acabo evaluaciones periódicas sobre sus actividades con el fin de planificar sus operaciones diarias y a mas largo plazo. Una forma de llevar acabo es mediante los llamados *Key Performance Indicators* (KPIs), que nos permiten medir el nivel de servicio, realizar diagnósticos de la situación, informar sobre los objetivos realizados.

Tomando como punto de partida los KPIs en el presente trabajo se propone analizar datos reales de vuelos realizados en USA , además se busca utilizar CML para identificar modelos que describan algunos comportamientos y evaluar su eficacia para realizar predicciones.

1.1 *Primer eje*

probando

1.2 *Segundo eje*

probando

2 Frontmatter

The biggest difference between a “usual” L^AT_EX style such as `article.sty` and the ENDM package is that the ENDM macro package requires the title, author’s name or names, abstract, keywords and “thanks” all to be included within the `frontmatter` environment. At the beginning of the source file for this paper, you’ll notice this. Also, you’ll notice that the usual `\maketitle` is absent; it no longer is needed. The ENDM style package automatically generates the title, author’s name and address, and related material at the beginning of the paper. Note also that `hyperref` has been disabled in this part of the `endm.cls` file, so references to footnotes aren’t linked to the appropriate footnotes or addresses. This is an old problem with L^AT_EX, involving the fact that the references within the frontmatter aren’t passed cleanly to the linking software.

For those who have used the ENDM package before, the one new thing to note is the inclusion of *Keywords* which are now required by Elsevier.

The ENDM macro package provides two alternatives to listing authors names and addresses. These are described in detail in the file `instraut.pdf`. Basically, listing each author and his or her address in turn, is the simplest method. But, if there are several authors and two or more share the same address (but not all authors are at this address), then the method of listing authors first, and then the addresses, and of referencing addresses to authors should be used.

Furthermore, note that an acknowledgment of support (the contents of `\thanks`) should be done by a separate listing of `\thanks[NSF]{To the NSF}` with the optional argument – `[NSF]` – being used for `\thanksref` which is attached to those authors acknowledging such support. It is important that the `\thanks` not be included within the scope of `\author{}` or of `\title{}`, but it must be within the scope of the environment `frontmatter`.

More details about added terms such as `\collab` can be found in the file `instraut.pdf` if they are needed.

3 Sectioning and Environments

Since ENDM is published through the auspices of Elsevier B. V., their style files were used to create the ENDM macro package. Below is a proof which shows that this package is not much different to most others:

Definition 3.1 A file is *derived* from another file if it was obtained by making only a few modifications to the original file.

Theorem 3.2 *The file `endm.cls` is derived from `elsart.sty`.*

Proof. This is clear from the similarity of the output to the output from the standard Elsevier style files. \square

If one wants to start a proof with a descriptive word, such as “sketch”, then one can use the `\begin{proof*}... \end{proof*}` environment, as in

Proof (Sketch) This can be derived from simple observations. \square

The main difference between the file `endm.cls` and the `elsart.cls` file used for other Elsevier journals is the more precise format we use. Elsevier’s generic style files are meant for preliminary editing and more precise formatting is imposed using a macro file designed for the specific Elsevier journal in which the paper will eventually appear. The `endm.cls` and `endmmacro.sty` files format papers uniformly so that they all are easily recognizable as belonging to the series *Electronic Notes in Discrete Mathematics*.

All of the usual features of L^AT_EX are available with these style files. It is only the formatting that has been rigorously defined. One can use the sectioning commands `\section`, `\subsection`, `\paragraph` and `\subparagraph`. The numbering scheme used is one under which Theorem 1.2.3 is the third numbered item in the second subsection of the first section of the paper. In order to facilitate cross-references, all of the named environments given below are numbered and all use the same numbering scheme.

The file `endmmacro.sty` contains additional information that is needed to typeset a paper. It also has the definitions of the *AMS* **euler** and **blackboard bold** fonts builtin. If you want to use symbols for the natural numbers, the reals, etc., then we prefer that you use the blackboard bold fonts, and not plain bold fonts. This is accomplished by using the `\mathbb` font, as in \mathbb{N} or \mathbb{R} .

The names of theorem-like environments are provided in `endmmacro.sty`. With the exception of the environment “Algorithm”, the names of all these are the full name rather than a shortened version. The environments provided

and their names are as follows:

- `\begin{theorem} ... \end{theorem}` for Theorems,
- `\begin{lemma} ... \end{lemma}` for Lemmas,
- `\begin{corollary} ... \end{corollary}` for Corollaries,
- `\begin{proposition} ... \end{proposition}` for Propositions,
- `\begin{criterion} ... \end{criterion}` for Criteria,
- `\begin{alg} ... \end{alg}` for Algorithms,
- `\begin{definition} ... \end{definition}` for Definitions,
- `\begin{conjecture} ... \end{conjecture}` for Conjectures,
- `\begin{example} ... \end{example}` for Examples,
- `\begin{problem} ... \end{problem}` for Problems,
- `\begin{remark} ... \end{remark}` for Remarks,
- `\begin{note} ... \end{note}` for Notes,
- `\begin{claim} ... \end{claim}` for Claims,
- `\begin{summary} ... \end{summary}` for Summary,
- `\begin{case} ... \end{case}` for Cases, and
- `\begin{ack} ... \end{ack}` for Acknowledgements.

For example,

Algorithm 1 *Step 1: Write the paper*

Step 2: Format it with the ENDM macro package

Step 3: Ship the whole thing to the Guest Editors

4 References and Cross-references

All the cross-referencing facilities of L^AT_EX are supported, so one can use `\ref{}` and `\cite{}` for cross-references within the paper and for references to bibliographic items. As is done in this note, the *References* section can be composed with `\begin{thebibliography}... \end{thebibliography}`. Alternatively, BibT_EX can be used to compile the bibliography. Whichever one is used, the references are to be numbered consecutively, rather than by author-defined acronyms. Of course you can use your own acronyms for easy reference to each of the items in the bibliography, as has been done with the listing for this short note.

Note that the references should *not* be started with a new `\section` command.

The package `hyperref` is automatically loaded by `endm.cls` and this makes all the cross-references within the document “active” when the pdf file of the paper is viewed with Adobe’s Acrobat[©] Reader. The format for including a link is simple: simply insert `\href{URL} {text}` where *URL* is the URL to which you want the link to point, and *text* is the text you want to be highlighted and which will bring up the desired web page when clicked upon.

4.1 Particulars about .pdf files

We now require that .pdf files be provided for publication online. A .pdf file is viewable by Adobe’s Acrobat[©] Reader, which can be configured to load automatically within a browser. Viewing a properly formatted .pdf file with Acrobat[©] allows the cross-references and links to URLs to be active. In fact, Elsevier utilizes .pdf files in order to take better advantage of the web’s capabilities.

One point that needs to be emphasized is that you should use Type 1 fonts when you typeset your L^AT_EX source file. These fonts are scalable, meaning that they carry information that allows the device viewing the final output to scale the fonts to suit the viewer being used (from an onscreen viewer such as Adobe’s Acrobat[©] Reader to printing the file on a printer). You can tell if you have used the right fonts by viewing the final output on your machine. If the font looks grainy, then you have not used a Type 1 font. Type 1 fonts can be located at the CTAN archive at <http://www.ctan.org>. They are public domain fonts and do not cost anything when you add them to your system.

Assuming you have Type 1 fonts available, there are several methods for producing .pdf files.

Using dvips and ps2pdf

We list this option first since it appears to be the most reliable and the easiest to use, especially if you include embedded PostScript graphics (.eps files) in your source file. Simply run L^AT_EX2e on your source file, apply `dvips` to produce a PostScript file and then apply `ps2pdf` to obtain a .pdf file.

The DVIPDFM utility

Another easy method for producing acceptable .pdf files is via the utility `dvipdfm`. This utility is included in distributions of MikT_EX, which runs on Windows machines, but it probably needs to be added to your teT_EX

distribution, if you are running L^AT_EX on a UNIX machine. The utility and precise information about installing it on your system can be found at the web page <http://gaspra.kettering.edu/dvipdfm/>. In essence, this utility converts a .dvi file into a .pdf file. So, one can first prepare the .dvi file using L^AT_EX, and then apply the utility `dvipdfm` to produce the needed .pdf file.¹ This utility makes the inclusion of graphics particularly simple. Those that are included in the L^AT_EX source file are simply converted to the .pdf format. As we note below, things are not so simple with the second alternative, which is to use pdfL^AT_EX.

pdfL^AT_EX

An alternative to the first possibilities to produce .pdf files is to process the source file with pdfL^AT_EX. This format is available from the standard CTAN sites <http://www.ctan.org>. It appears that pdfL^AT_EX and `hyperref` have some problems when used together. It is necessary to use pdfL^AT_EX version 14d or later in order to minimize these issues. If your system has an earlier version (most t_EX distributions have version 13d), then you can update your system by retrieving the latest version of pdfL^AT_EX from <ftp://ftp.cstug.cz/pub/tex/local/cstug/thanh/pdftex/>. Even if the recent versions are used, pdfL^AT_EX has the same dealing with references embedded with the `frontmatter` section described above for L^AT_EX.

But there is one aspect of pdfL^AT_EX that creates problems. Many authors include EPS² files within their papers. While this is fairly straightforward with L^AT_EX, there are a couple of points to note when attempting this with pdfL^AT_EX.

To include a PostScript image in a .pdf file produced with pdfL^AT_EX, you first have to convert the image to a .pdf file. The conversion can be accomplished most easily using Ghostscript; you can simply view the file in Ghostview and then print the image to a .pdf file using the `pdfwriter` option within Ghostview. The result for a standard chess board that is part of the Ghostview distribution is the following image:

Below is a copy of a color image. While pdfL^AT_EX can handle image files in other formats, L^AT_EX can only handle .eps images reliably.

¹ *Beware!* The utility `dvipdf` does *not* produce acceptable .pdf files, and should not be used. Only `dvipdfm` should be used to produce .pdf files.

² EPS stands for *embedded PostScript*, which affords a mechanism for including pre-prepared PostScript files within a L^AT_EX document.

Using ENDM Macros with Mac OS X

Clearly, if your file does not require `.eps` or other PostScript files, then you can create the required `.pdf` file using any of the standard \TeX implementations for the Macintosh. If you do need to include PostScript files and if you are using \TeX Shop, then you can specify to use `dvips` and Ghostview in processing your file, and then you can apply `ps2pdf` to create the needed `.pdf` file. Alternatively, the Mac OS X operating system is based on UNIX, so it supports the use of te\TeX as described above.

5 Summary and Remarks

The ENDM macro package is relatively easy to use and provides a uniform layout for all the papers that appear in ENDM.

Assigning Volume Numbers

An additional point worth mentioning is that ENDM has moved to *ScienceDirect*, Elsevier's main platform for publishing electronic series. Because *ScienceDirect* cannot easily accommodate changes to published material, the *Proceedings* must be entirely ready before they can be published. Volume numbers will therefore not be assigned for the *Proceedings* until the final versions of all papers are in.

Copyright Transfer Forms

Due to the move to *ScienceDirect*, the corresponding author of each paper published in ENDM must submit a signed Copyright Transfer Form to Elsevier in order for their paper to be published. A copy of this form will be sent to each author. Note that the publication of an abstract or extended abstract in ENDM will not restrict the author(s) from publishing a full-length article on the same topic and with the same title in another journal (possibly with another publisher). Details about the copyright agreement specifying the exact rights of the authors and the rights of Elsevier are available at [Elsevier's Author Gateway](#).

6 Bibliographical references

ENDM employs the `plain` style of bibliographic references in which references are numbered sequentially and listed in alphabetical order according to the first author's last name. Please utilize this style. We have a BibTeX style file, for those who wish to use it. It is the file `endm.bst` which is included in this package. The basic rules we have employed are the following:

- Authors' names should be listed in alphabetical order, with the first author's last name listed first followed by initials or first name, and with the other authors' names listed as *first name, last name*.
- Titles of articles in journals should be in *emphasized* font.
- Titles of books, monographs, etc. should be in quotations.
- Journal names should be in plain roman type.
- Journal volume numbers should be in boldface, immediately followed by the year of publication enclosed in parentheses in roman type.
- References to URLs on the net should be "active" and the URL itself should be in `typewriter` font.
- Articles should include page numbers.

The criteria are illustrated by the examples below.

References

- [1] Civin, P., and B. Yood, *Involutions on Banach algebras*, Pacific J. Math. **9** (1959), 415–436.
- [2] Clifford, A. H., and G. B. Preston, "The Algebraic Theory of Semigroups," Math. Surveys **7**, Amer. Math. Soc., Providence, R.I., 1961.
- [3] Freyd, Peter, Peter O'Hearn, John Power, Robert Tennent and Makoto Takeyama, *Bireflectivity*, Electronic Notes in Theoretical Computer Science **1** (1995), URL: <http://www.elsevier.com/locate/entcs/volume1.html>.
- [4] Easdown, D., and W. D. Munn, *Trace functions on inverse semigroup algebras*, U. of Glasgow, Dept. of Math., preprint 93/52.
- [5] Roscoe, A. W., "The Theory and Practice of Concurrency," Prentice Hall Series in Computer Science, Prentice Hall Publishers, London, New York (1198), 565pp. With associated web site <http://www.comlab.ox.ac.uk/oucl/publications/books/concurrency/>.

- [6] Shehadah, A. A., "Embedding theorems for semigroups with involution, " Ph.D. thesis, Purdue University, Indiana, 1982.
- [7] Weyl, H., "The Classical Groups," 2nd Ed., Princeton U. Press, Princeton, N.J., 1946.