



Universidad Nacional de Rosario
Secretaría de Ciencia y Tecnología

**INFORME FINAL
DE PROYECTOS DE INVESTIGACIÓN
FORMULARIO DE PRESENTACIÓN**

1.IDENTIFICACIÓN DEL PROYECTO

1.1.DENOMINACIÓN DEL PROYECTO

1AGR145 ANALISIS ESTADISTICO MULTIVARIADO EN EL ESTUDIO Y COMPARACION DE DISTINTOS TIPOS DE TEXTOS ACADEMICOS: CIENCIAS BIOLOGICAS Y CIENCIAS SOCIALES

AÑO DE INICIO: 2010

AÑO DE FINALIZACIÓN: 2013

1.2.DIRECTOR DEL PROYECTO

DIRECTOR

Apellido y Nombre: Beltrán, Celina

CUIL : 27 - 21722712 - 2

Domicilio particular: Primero de mayo 1464- Casilda

Domicilio laboral: Campo Experimental Villarino cc14- Zavalla

Teléfono: 03464-425164

FAX:

E-mail: beltranc@dat1.net.ar

1.3.RADICACIÓN DEL PROYECTO

DEPENDENCIA :FAC. DE CS. AGRARIAS

UNIDAD EJECUTORA:Facultad de Ciencias Agrarias

2.LOGROS DEL PROYECTO*

*No es necesario informar sobre todos los items sólo de aquellos en los que se hayan producido logros.

2.1. LOGRO DE LOS OBJETIVOS DEL PROYECTO (contribución al avance del conocimiento científico y tecnológico)

* Reproduzca aquí los objetivos originalmente planteados y luego describa el cumplimiento de los mismos.

OBJETIVOS PLANTEADOS EN EL PROYECTO

-Utilizar las técnicas estadísticas multivariadas para describir y eventualmente comparar las estructuras de distintos tipos de textos académicos provenientes de distintas áreas científicas: Ciencias Biológicas y Ciencias Sociales.

-Realizar el análisis automático de textos científicos provenientes de distintas áreas científicas: Ciencias Biológicas y Ciencias Sociales.

SU CUMPLIMIENTO

Respecto a los datos:

-Se realizó el análisis automático de 180 textos académicos.

-Se logró confeccionar una base de datos formada por 180 textos conteniendo la información procedente del análisis morfológico de éstos.

-La arquitectura de la base de datos fue “ajustada” o modificada cuando así fue requerido de modo de poder aplicar las técnicas estadísticas pertinentes de manera adecuada. En algunos casos fue necesaria la transformación de las variables originales ampliando así la cantidad de campos de la base.

-Durante el proceso de estimación de los modelos se confeccionaron los programas necesarios para ejecutar las técnicas de clasificación en distintos software estadísticos y se comparó la información brindada por cada uno de ellos en cada tipo de análisis.

Respecto a los resultados encontrados en la clasificación según la disciplina de los textos se pueden enumerar los siguientes resultados:

-Los resultados del análisis morfológico de los textos se analizaron teniendo en cuenta simultáneamente todas las mediciones realizadas sobre ellos.

-El análisis de regresión logística aplicado presenta una modalidad de análisis estadístico para discriminar grupos no muy habitual en la investigación lingüística y permitió hallar las categorías gramaticales cuyas frecuencias observadas en los textos permiten discriminar los grupos definidos por la disciplina a la que pertenecen. Las diferencias entre los textos de BIOMETRÍA y FILOSOFÍA está centrada principalmente en el número de clíticos y de adverbios presentes. Los odds ratio estimados evidencian que la chance de clasificar a un texto dentro del corpus de Biometría se incrementa en un 62% al aumentar en número de clíticos en una unidad, mientras que la chance de clasificarlo en el corpus de Filosofía aumenta un 41% al incrementarse en una unidad el número de adverbios. Similares resultados se hallaron utilizando un análisis discriminante sobre las variables transformadas. Este resultado puede deberse a que, en los textos de biometría/estadística hay más clíticos que en los humanísticos por la frecuencia de expresiones impersonales o pasivas con el clítico “se” del tipo: “se ajusta un modelo cuadrático”, “se estima la variancia poblacional”; mientras en los textos de filosofía se manifiesta la presencia de mayor proporción de adverbios.

-Respecto a la metodología basada en Árboles de Clasificación, si bien el número de unidades utilizadas en el entrenamiento y evaluación no era elevado, el árbol obtenido mostró un buen desempeño frente al modelo de regresión logística. La diferencia en la tasa de mala clasificación sólo se diferenció en el área de Filosofía.

-En los clasificadores basados en Regresión Logística y en el Análisis Discriminante, las diferencias entre los dos tipos de textos se basa principalmente en el porcentaje de clíticos y de adverbios presentes. Sin embargo, en el clasificador basado en Árboles de Clasificación han intervenido otras variables en la separación como el porcentaje de preposiciones y adjetivos. Estas variables intervienen determinando una interacción entre las categorías gramaticales que no se alcanza a observar en el modelo de regresión logística.

-La técnica de Regresión Logística se generalizó a un número mayor de disciplinas, de las cuales provienen los textos, mediante una extensión del modelo para variable respuesta multinomial. El análisis de Regresión Logística Multinomial aplicado permitió hallar las categorías gramaticales cuyas frecuencias observadas en los textos permiten discriminar los tres grupos definidos.

2.2. LOGROS METODOLÓGICOS

La utilización que se ha hecho de estas técnicas multivariadas en la caracterización de textos académicos considerando la distribución de las categorías gramaticales de los mismos, es una aplicación novedosa en la lingüística computacional. Esto se debe principalmente que se le ha dado la importancia adecuada a la interpretación de los coeficientes de los modelos estadísticos logrando una descripción muy clara de los textos en función de la información proveniente del análisis morfológico automático.

2.3. CONTRIBUCIÓN A LA FORMACIÓN DE RECURSOS HUMANOS

Dos miembros del equipo becarios del CONICET han terminado durante la ejecución del proyecto el doctorado en la Facultad de Humanidades y Artes de la UNR sobre temas de Lingüística Computacional. Otro miembro, durante la ejecución del proyecto, ha iniciado y finalizado sus estudios en la maestría de estadística Aplicada de la Facultad de Ciencias Económicas de la UNR.

2.4.CONTRIBUCIÓN AL DESARROLLO ECONÓMICO Y SOCIAL

La clasificación automática de textos ha recibido más atención en los últimos tiempos debido a la gran cantidad de información disponible en formato electrónico. Esta nueva perspectiva en el estudio de las técnicas estadísticas multivariadas en la clasificación de textos académicos significa una contribución de importancia y una contribución novedosa.

2.5.TRANSFERENCIA DE RESULTADOS REALIZADAS

Se han realizado publicaciones, seminarios y presentaciones a congresos..

Los resultados que se refieren a la aplicación de la metodología estadística y a la confección de la base de datos han sido transferidos a los cursos de posgrado que he dictado en el período correspondiente al informe:

-Metodología de la Investigación, Doctorado de Humanidades y Artes, Mención Lingüística. Universidad Nacional de Rosario. Docentes: Dra. Celina Beltrán- Dr. Diego Beltrán. Años 2011, 2012 y 2013. Duración: 30 hs.

-Estadística I. Doctorado en Ciencias Agrarias. Facultad de Ciencias Agrarias. Universidad Nacional de Rosario. Años 2011,2012 y 2013. Duración: 50 hs.

-Estadística I. Maestría en Genética Vegetal. Facultad de Ciencias Agrarias. Universidad Nacional de Rosario. Años 2011,2012 y 2013. Duración: 50 hs.

-Estadística I. Maestría en Manejo y Conservación de Recursos Naturales. Facultad de Ciencias Agrarias. Universidad Nacional de Rosario. Años 2011, 2012 y 2013. Duración: 50 hs.

-Introducción al Análisis Multivariado. Maestría en Genética Vegetal. Facultad de Ciencias Agrarias. Universidad nacional de Rosario. Años 2010 y 2012. Duración 40 hs.

2.6.PERSPECTIVAS DE FUTURA TRANSFERENCIA

Se espera profundizar la transferencia iniciada.

2.7. DIVULGACIÓN REALIZADA (Publicaciones, comunicaciones, etc)

Las siguientes publicaciones han surgido de esta investigación:

Libros:

“Introducción al análisis estadístico en la investigación. Con aplicaciones en distintas disciplinas” Autor: Celina Beltrán. Ediciones Juglaría. Rosario. Año 2012. I.S.B.N. 978-987-1166-63-3.

Capítulos de libros:

Capítulo de libro (Cap. 2): “Algunos aspectos del análisis estadístico en la investigación” Nombre del libro: “Metodología de la Investigación. Teoría, práctica y ejemplos”. GRUPO INFOSUR- Ediciones Juglaría. I.S.B.N. 978-987-1166-54-1. Autor del Capítulo: Beltrán, Celina. Año 2011.

Artículos en revistas:

-“Estudio y comparación de distintos tipos de textos académicos: Biometría y Filosofía”. Autor: Celina Beltrán. Revista de Epistemología y Ciencias Humanas. Nro. 2 Año 2010. ISSN 1852-625X.

- “Análisis discriminante aplicado a textos académicos: Biometría y Filosofía” Autor: Celina Beltrán. Revista INFOSUR. ISSN: 1851-1996. Número 4. Año 2010.

-“Análisis de regresión logística aplicado a la clasificación textos académicos: Biometría y Filosofía”. Autor: Celina Beltrán. Revista de Epistemología y Ciencias Humanas. Nro. 3 Año 2011. ISSN 1852-625X.

-“Aplicación del análisis de regresión logística multinomial en la clasificación de textos académicos: Biometría, Filosofía y Lingüística informática” Autor: Celina Beltrán. Revista INFOSUR ISSN: 1851-1996. Número 5. Año 2011.

-Extracción del Sintagma Nominal. Autores: Solana Zulema, Rodrigo Andrea, Méndez Bárbara. Revista INFOSUR. ISSN: 1851-1996. Número 5. Año 2011.

-“Aplicación de redes neuronales artificiales en la clasificación de textos académicos según disciplina: Biometría, Filosofía y Lingüística informática”. Autor: Celina Beltrán. Revista de Epistemología y Ciencias Humanas. Nro. 4 Año 2012. ISSN 1852-625X.

-Árboles de clasificación y su comparación con análisis de regresión logística aplicado a la clasificación de textos académicos”. Autor: Celina Beltrán. Revista INFOSUR. ISSN: 1851-1996. Número 6. Año 2012

-Los sistemas de ayuda en la navegación hipertextual. Autor: Bárbara Méndez. Revista INFOSUR. ISSN: 1851-1996. Número 6. Año 2012

-Estudio exploratorio para la comparación de distintos tipos de textos: Textos Científicos y Textos No Científicos. Autor: Celina Beltrán. Revista de Epistemología y Ciencias Humanas. Nro. 5 Año 2013. ISSN 1852-625X.
 - ANÁLISIS AUTOMÁTICO MORFOLÓGICO CON LAS HERRAMIENTAS SMORPH Y NOOJ. Autor: Celina Beltrán, Zulema Solana, Carolina Tramallino. Revista de Epistemología y Ciencias Humanas. Nro. 5 Año 2013. ISSN 1852-625X.

Trabajos presentados a congresos:

-La sufijación nominal en la interlengua de los aprendientes de español como L2: análisis automático e índice de sufijación. Autores: Celina Beltrán, Zulema Solana. XII Congreso de la Sociedad Argentina de Lingüística. 6 al 9 de abril de 2010. Mendoza.

-“Reconocimiento automático de la interlengua: nombres con marcas flexivas de género y número diferentes a la lengua estándar.” Autor: MÉNDEZ, BÁRBARA.. XII Congreso de la Sociedad Argentina de Lingüística: La renovación de la palabra, Facultad de Filosofía y Letras, UNCUYO, abril 2010,

-Resolución de ambigüedades en el etiquetado morfológico de un texto mediante un modelo de regresión logística y su comparación con reglas lingüísticas. Autor: Celina Beltrán. XXXIX Coloquio Argentino de Estadística. 12 al 14 de Octubre 2011.

-Redes neuronales y regresión logística en la clasificación automática de textos. Autores: Celina Beltran, Alberto Trevizan. X Congreso Latinoamericano de Sociedades de Estadística. Córdoba, Argentina. 16 a 19 de octubre 2012.

3. PRESUPUESTO EJECUTADO

(**)	ORIGEN DEL FINANCIAMIENTO(*)	PRESUPUESTO EJECUTADO EN EL AÑO 2010 (en \$)	PRESUPUESTO EJECUTADO EN EL AÑO 2011 (en \$)	PRESUPUESTO EJECUTADO EN EL AÑO 2012 (en \$)	PRESUPUESTO EJECUTADO EN EL AÑO 2013 (en \$)
7164	U.N.R	3695	3469	0	0

(*): Excluidos los salarios y/o becas de los docentes-investigadores

(**): Puede consignarse monto global o discriminado en rubros

Nota: Si los años de duración del proyecto son mas de 3, repetir el cuadro precedente.

4. RECURSOS HUMANOS

4.1. INTEGRACIÓN DEL EQUIPO DE TRABAJO

APELLIDO Y NOMBRE	PERIODO EN EL QUE PARTICIPO EN LA EJECUCIÓN DEL PROYECTO	FUNCIÓN DENTRO DEL PROYECTO	MÁXIMO TÍTULO ACADÉMICO ALCANZADO	CARGO DOCENTE	DEDICACION	CARGO CIUNR	CARGO CONIC ET	HS. SEM. DED. AL PROY.	CATEGORÍA FIRME EN EL PROG. DE INCENTIVOS	CATEGORÍA EN TRÁMITE EN EL PROG. DE INCENTIVOS	UNIVERSIDAD	FIRMA
BELTRAN, CELINA	2010-2013	DIRECTOR	pos doctorado	J.T.P.	Exclusiva	Ninguno	Ninguno	22	III	Ninguna	UNR	
FERNANDEZ DIMATTEO, GABRIELA ALEJANDRA	2010-2013	INTEGRANTE	profesora	Prof. ADJUN	Exclusiva	Ninguno	Ninguno	10	V	Ninguna	UNR	
Koza, Walter	2010-2013	INTEGRANTE	doctorado	Aux. 1era.	Simple	Ninguno	Becario	6	V	Ninguna	UNR	
Méndez, Bárbara	2010-2013	INTEGRANTE	doctorado	Aux. 1era.	Simple	Ninguno	Ninguno	6	Ninguna	Ninguna	UNR	
Freidenberg, Susana	2011-2013	INTEGRANTE	magister	Prof. ADJUN	Exclusiva	Cat. C	Ninguno	12	III	Ninguna	UNR	
Rivero, Silvia	2011-2013	INTEGRANTE	magister	Prof. ADJUN	Simple	Ninguno	Ninguno	12	III	Ninguna	UNR	
		DIRECTOR		Ninguno	Ninguna	Ninguno	Ninguno		Ninguna	Ninguna	UNR	
		DIRECTOR		Ninguno	Ninguna	Ninguno	Ninguno		Ninguna	Ninguna	UNR	
		DIRECTOR		Ninguno	Ninguna	Ninguno	Ninguno		Ninguna	Ninguna	UNR	
		DIRECTOR		Ninguno	Ninguna	Ninguno	Ninguno		Ninguna	Ninguna	UNR	
		DIRECTOR		Ninguno	Ninguna	Ninguno	Ninguno		Ninguna	Ninguna	UNR	
		DIRECTOR		Ninguno	Ninguna	Ninguno	Ninguno		Ninguna	Ninguna	UNR	
		DIRECTOR		Ninguno	Ninguna	Ninguno	Ninguno		Ninguna	Ninguna	UNR	
		DIRECTOR		Ninguno	Ninguna	Ninguno	Ninguno		Ninguna	Ninguna	UNR	
		DIRECTOR		Ninguno	Ninguna	Ninguno	Ninguno		Ninguna	Ninguna	UNR	
		DIRECTOR		Ninguno	Ninguna	Ninguno	Ninguno		Ninguna	Ninguna	UNR	

4.2. OTROS RECURSOS HUMANOS QUE PARTICIPARON EN EL EJECUCIÓN DEL PROYECTO

APELLIDO Y NOMBRE	DNI	FUNCIÓN DENTRO DEL PROYECTO	PERÍODO EN EL QUE PARTICIPO	MÁXIMO TÍTULO ACADÉMICO ALCANZADO	HS. SEM. DED. AL PROY.	FIRMA

Nota: Completar el cuadro precedente con los datos de alumnos, becarios, pasantes y graduados que hayan integrado el equipo, de acuerdo al marco regulatorio de cada facultad.

5. REGISTRO INSTITUCIONAL

La Unidad Académica donde se radicó el Proyecto hace constar el registro del presente informe.

Firma :

Aclaración :

Cargo :