

Plan de Gestión de Datos

INFORMACION SOBRE EL PROYECTO	
1. – Título del Proyecto	
- Título del Proyecto (en castellano)	
Desarrollo de métodos de aprendizaje estadístico y su aplicación a economía y ciencias sociales.	
- Título del Proyecto (en ingles)	
Development of statistical learning methods and their application to economics and social sciences.	
-Descripción del Proyecto (en castellano) Resumen	
Uno de los principales objetivos en el ámbito de la estadística es la capacidad de predecir fenómenos de interés a partir de ciertas variables observadas. Cuando se construye un modelo de predicción es importante tener en cuenta la naturaleza de la variable a predecir y las características de las variables con las que se cuenta para realizar la estimación del modelo. Los modelos clásicos se basan en ciertos supuestos sobre las variables involucradas, como por ejemplo independencia y distribución de las mismas. Es sabido que modelos desarrollados para datos independientes no tienen un buen desempeño en datos que presentan cierta estructura de dependencia. Por otra parte, los avances tecnológicos que venimos experimentando en la era del Big Data facilitan la generación de datos que a menudo no está acompañada por un incremento similar en la cantidad de observaciones recolectadas. Cuando los métodos usuales de estimación son aplicados a los casos antes descriptos, las estimaciones proporcionadas por los mismos son poco fiables y sesgadas. Para sortear estas limitaciones es de interés proponer modelos que se adecúen a los datos disponibles y a los objetivos buscados, particularmente, a datos de alta dimensión, de conteo con exceso de ceros y/o datos espaciales que naturalmente presentan dependencia. Este tipo de datos es muy frecuente en el área de economía y ciencias sociales, por lo que en el presente proyecto nos focalizaremos en tales aplicaciones.	
-Descripción del Proyecto (en ingles) Resumen	
One of the main objectives in the field of statistics is the ability to predict phenomena of interest from certain observed variables. When building a prediction model, it is important to consider the nature of the variable to be predicted and the characteristics of the variables available for estimating the model. Classical models are based on certain assumptions about the variables involved, such as independence and distribution. It is well known that models developed for independent data do not perform well on data that exhibit some structure of dependence. Moreover, the technological advances we have been experiencing in the Big Data era facilitate the	

generation of data that is often not accompanied by a similar increase in the number of observations collected. When usual estimation methods are applied to the aforementioned cases, the estimates provided by them are unreliable and biased. To overcome these limitations, it is of interest to propose models that are suited to the available data and the desired objectives, particularly high-dimensional data, count data with excess zeros, and/or spatial data that naturally present dependence. This type of data is very common in the fields of economics and social sciences, so in the present project, we will focus on such applications.

-Palabras Clave descriptivas del Proyecto (en castellano)

Modelos lineales generalizados;
Modelos semiparamétricos;
Datos espacialmente correlacionados;
Modelos censurados/inflados en cero;
Estimadores robustos;
Reducción de dimensiones

- Palabras Clave descriptivas del Proyecto (en ingles)

Generalized linear models;
Semiparametric models;
Spatially correlated data;
Censored and zero inflated models;
Robust estimators;
Dimension reduction

2 – Datos del Director/ar del Proyecto

- Nombre y Apellido Pamela Llop

- Unidad Académica Facultad de Ingeniería Química

- Teléfono oficial de contacto 4571164 int 314

-Teléfono móvil de contacto 3425170573

-E-mail del Director/a del Proyecto lloppamela@gmail.com

DATOS RESULTANTES DE LA EJECUCIÓN DEL PROYECTO

-Describa la toma de muestras / datos a realizar

En este proyecto se utilizarán datos ya recolectados, bases libres y encuestas de hogares extraídas de Institutos oficiales de Estadística (INDEC, IPEC, CONI).

– Datos: ¿Existe alguna razón por la cual los datos declarados no deban ser puestos a disposición de la comunidad / ser de acceso público? (marque X)

☒ **NO**

SI. Elija una de las opciones:	
	a) se encuentra en evaluación de protección por medio de patentes b) no se inició el proceso de evaluación de patentabilidad, pero podría ser protegible c) existe un contrato con un tercero que impide la divulgación d) Otro. Justifique.
– Período de Confidencialidad: Es el periodo durante el cual los datos no deberían ser publicados, contado a partir del momento de la toma de los mismos. El periodo máximo para la no publicación es de 5 (CINCO) años posteriores a su obtención. Luego de este periodo, los datos estarán disponibles para la comunidad / serán de acceso público. Si Ud. considera que este tiempo es insuficiente, y necesita prorrogar el período de confidencialidad, indique sus motivos y la cantidad de años adicionales que considera necesarios. Marque su opción con “X”.	
	1 (UN) año
	2 (DOS) años
	3 (TRES) años
	4 (CUATRO) años
	5 (CINCO) años
X	Otro.
Motivos: Los datos son de acceso público	

INSTRUCTIVO PARA LLENADO DEL PLAN DE GESTIÓN DE DATOS

El PGD no es un documento definitivo, sino que se desarrollará a lo largo del ciclo de vida del proyecto.

INFORMACIÓN SOBRE EL PROYECTO

1- Título del Proyecto (en castellano): Deberá ingresar el título completo del proyecto (en castellano), indicando además el código asignado por la SCAYT.

- Título del Proyecto (en inglés): Deberá ingresar el título completo del proyecto en inglés.
- Descripción del Proyecto (en castellano): Deberá ingresar la descripción del Proyecto en castellano.
- Descripción del Proyecto (en inglés): Deberá ingresar la descripción del Proyecto en inglés.
- Palabras Clave descriptivas del Proyecto (en castellano): Deberá ingresar tres palabras clave descriptivas del Proyecto, en castellano.
- Palabras Clave descriptivas del Proyecto (en inglés): Deberá ingresar tres palabras clave descriptivas del Proyecto, en inglés.

2- Datos del Director/a del Proyecto

- Nombre y Apellido del Titular del Proyecto: Nombre completo y apellido del Titular del Proyecto.
- Unidad Académica: Nombre de la UA a la que pertenece el /la directora/a del Proyecto.
- Teléfono oficial de contacto: Número de teléfono de la oficina / laboratorio / Institución del Director/a del Proyecto, donde pueda ser contactado, incluyendo número de área / país (ej: Para la Santa Fe: + 54 9 342 4999-9999).
- Teléfono móvil de contacto: Número de t
- E-mail del Director/a del Proyecto: Correo electrónico de contacto del Director/a del Proyecto.

DATOS RESULTANTES DE LA EJECUCIÓN DEL PROYECTO

- Describa la toma de muestras / datos a realizar: Información descriptiva sobre la toma de muestras que resultaran en datos / conjuntos de datos. La descripción deberá incluir información de contexto (lugar de toman los datos; instrumentos etc).

Datos: ¿Existe alguna razón por la cual los datos declarados no deban ser puestos a disposición de la comunidad / ser de acceso público? Deberá marcar con una “X” la opción correcta. En caso de responder afirmativamente, deberá justificar debidamente, comprendiendo que solo en casos de extrema excepcionalidad esta restricción de acceso a los datos resulta practicable / aceptable.

-Período de Confidencialidad: Es el periodo durante el cual los datos no deberían ser publicados, contado a partir del momento de la toma de los mismos. El periodo máximo para la no publicación es de 5 (CINCO) años posteriores a su obtención. Luego de este periodo, los datos estarán disponibles para la comunidad / serán de acceso público. Si Ud. considera que este tiempo es insuficiente, y necesita prorrogar el período de confidencialidad, indique sus motivos y la cantidad de años adicionales que considera necesarios.