**Carrera:** Ingeniería en Informática.

Universidad Católica de

Santiago del Estero

**Asignatura:** Trabajo Final.

**Título:** Análisis de datos para resolución de cuestiones estratégicas relacionadas con la lechería en el INTA.

**Tema:** Plan de Proyecto.

**Alumnos:**

* Pellegrinet, Nicolás.

**Ciudad:** Rafaela.

***Contenido***

[Descripción del problema 1](#_Toc491777677)

[Objetivos y metas 2](#_Toc491777678)

[Antecedentes 2](#_Toc491777679)

[Justificación 3](#_Toc491777680)

[Marco teórico 4](#_Toc491777681)

[Marco metodológico 6](#_Toc491777682)

[Tipo de investigación 6](#_Toc491777683)

[Técnicas de recolección de datos 7](#_Toc491777684)

[Plan de tratamiento de datos 7](#_Toc491777685)

[Alcance 8](#_Toc491777686)

[Cronograma 9](#_Toc491777687)

[Bibliografía 10](#_Toc491777688)

# Descripción del problema

Tal como se menciona en el sitio web (Presentamos el tambo robotizado, 2015) del Instituto Nacional de Tecnología Agropecuaria (INTA), la sede del instituto ubicada en la ciudad de Rafaela, dispone de la instalación de un tambo robotizado. El mismo permite que el proceso de ordeño y alimentación sea totalmente automatizado, denominando a esta modalidad como “Sistema de Ordeño Voluntario” (VMS) o también “Sistema de Ordeño Automático” (por sus siglas en inglés AMS). Se trata del primer tambo robotizado del país y el primero de Latinoamérica vinculado con una institución.

La instalación de este tambo, se llevó a cabo gracias al *Convenio de Vinculación Tecnólógica entre el INTA (Centro Regiona Santa Fe – EEA Rafaela) y la empresa DeLaval.* El principal objetivo de este convenio es lograr adaptar este tipo de tecnología en un sistema de producción representativo de la Argentina.

El proyecto comenzó en octubre de 2014 y, en general, fue divido en tres etapas:

* Primera etapa: obra civil (con apoyo económico de la Asociación Cooperadora de la EEA Rafaela del INTA.
* Segunda etapa: selección de personal.
* Tercera etapa: incremento del rodeo a 60-70 vacas en ordeño y su estabilización.

Adicionalmente, este tambo cuenta con un sistema computarizado que almacena gran cantidad de información asociada. Si bien los datos son guardados y procesados por un programa que devuelve información útil para la gestión y toma de decisiones, se considera que esta información no está siendo aprovechada en su totalidad.

De este modo, se plantea como solución la realización de un análisis exploratorio de los datos de forma tal de resolver cuestiones estratégicas para la actividad de lechería. Para lograr dicho propósito, la utilización de técnicas de procesamiento de datos e inteligencia artificial, puede permitir la elaboración de curvas matemáticas asociadas y la detección de patrones que permitan mejorar el proceso de toma de decisiones.

# Objetivos y metas

Objetivo general

El proyecto tiene como objetivo contribuir con la investigación que está llevando a cabo el INTA en relación al tambo robotizado. De esta manera, se plantea realizar un procesamiento y análisis exploratorio de los datos para obtener así gráficas e indicadores que permitan tomar decisiones acerca de las prácticas asociadas a la actividad lechera.

Objetivos específicos

* Realizar un correcto análisis y tratamiento de la información.
* Utilizar técnicas y herramientas adecuadas para un análisis exploratorio de los datos.
* Detectar y eliminar anomalías y errores contenidos dentro del set de datos.
* Presentar un documento parametrizado con información útil y relevante.
* Obtener gráficas y curvas matemáticas en función de la información disponible.

# Antecedentes

El mismo sitio (Presentamos el tambo robotizado, 2015), menciona que la aplicación de la robótica a la actividad lechera es una cuestión que lleva más de 20 años de desarrollo. El principal desafío es lograr adaptar este tipo de tecnología en tambos pastoriles intensivos y de mayor complejidad.

Sin embargo, existen una gran cantidad de investigaciones derivadas de la instalación de este tambo. Según explicitó el personal del INTA, algunas de estas investigaciones no tienen antecedentes en el mundo.

De esta manera, con el presente trabajo se facilitarán herramientas e indicadores útiles para obtener conclusiones generales y específicas de cada una las actividades que componen el proceso de lechería.

# Justificación

El presente trabajo surge gracias al convenio establecido entre el INTA y la Universidad Católica de Santiago del Estero (Departamento Académico Rafaela).

Desde el INTA se planteó la necesidad de analizar y explorar los datos almacenados por el sistema anteriormente descripto, para obtener así información que permita responder a ciertos interrogantes estratégicos asociados a la actividad lechera.

Actualmente, muchos de los experimentos realizados tanto por el INTA como por otras entidades, tienen la particularidad de que se llevan a cabo con una muestra muy reducida de vacas. Estos animales son ordeñados bajo ciertas modificaciones en el proceso y, luego de ser analizados durante meses, se elaboran conclusiones acerca de las hipótesis planteadas en primera instancia.

Con los datos almacenados por el sistema del tambo robotizado, se cuenta con un registro histórico de cada ordeño de un gran número de vacas. De esta manera, se tiene una población de estudio mucho más amplia que la de cualquier otro experimento realizado anteriormente.

Así, el presente proyecto de investigación puede contribuir con el descubrimiento de grandes aportes y conclusiones para la actividad.

# Marco teórico

Para comenzar, se debe establecer claramente qué es un ordeño. De acuerdo con la definición de la Real Academia Española (RAE), “ordeño es extraer la leche exprimiendo una ubre”. Haciendo un análisis de esta definición y elaborando un concepto más específico, se puede decir que el ordeño es el procedimiento mediante el cual se extrae la leche de las glándulas mamarias de un mamífero.

Anteriormente ya se mencionó que el presente trabajo hace alusión al ordeño de vacas. Por este motivo, tal como se detalla el documento publicado por el INTA (Bretschneider, Salado, Cuatrin, & Arias) existen una serie de conceptos asociados a este tipo de animal y que deben ser considerados para comprender la totalidad de la presente investigación.

Lactancia

El termino lactancia hace referencia a la generación de leche por parte de la vaca. Este “proceso”, se origina con el parto y por ende depende de forma exclusiva de la gestación. Es importante mencionar que, para ganar productividad, la vaca es preñada mientras está en producción.

Período seco

Puesto que la vaca es preñada mientras se encuentra produciendo leche, la misma es “secada”, en general, dos meses previos al parto. Dicho de otro modo, la vaca deja de ser ordeñada dos meses previos al parto.

Ciclo productivo

El ciclo productivo de una vaca es el período de tiempo mediante el cual el animal produce leche y tiene una duración aproximada de 305 días.

Número de lactación

Cada ciclo productivo de una vaca es contabilizado de manera acumulativa a lo largo de su vida. De esta manera, el número de lactación, hace referencia a la cantidad de ciclos productivos que ha tenido una vaca.

Pico de lactancia

Es el nivel más alto de producción que una vaca puede alcanzar dentro de los primeros 90 días de lactación (luego del parto). De acuerdo con en el documento elaborado por el INTA (EEA Rafaela), “existe una relación positiva entre el pico y la subsecuente producción de leche a lo largo de la lactancia. Dicho de otra manera, a medida que los litros de leche al pico incrementan, también incrementan los litros totales producidos por lactancia”

Tasa de descenso

Luego de alcanzar el pico de lactancia la producción comienza a descender gradualmente hasta llegar al final la lactancia. Esto es lo que se conoce como tasa de descenso.

Persistencia

Este término hace referencia a la tasa de descenso en la producción luego de haber alcanzado el pico de lactancia. Es decir, a mayor tasa de descenso habrá una menor persistencia mientras que, cuanto mayor sea la persistencia, menor será su tasa de descenso.

Intervalos de ordeño

En la actividad de lechería, se conoce como intervalo de ordeño, a la diferencia de tiempo existente entre el ordeño de un animal y el ordeño inmediatamente anterior. A modo de ejemplo, si una vaca es ordeñada a las 18 hs. y su ordeño anterior sucedió a las 12 hs. entonces el intervalo para el primer ordeño (realizado a las 18 hs.) es de 6 hs.

# Marco metodológico

Según los conceptos descriptos por los autores Roberto Hernández Sampieri, Carlos Fernández Collado y Pilar Baptista Lucio (2006), se procede con el desarrollo del marco metodológico.

## Tipo de investigación

Según el paradigma, la presente investigación es considerada ***interpretativa*** ya que se enfoca plenamente en la comprensión del problema planteado y en obtener verdaderos significados que contribuyan con la investigación llevada a cabo por el INTA.

Por otro lado, al realizar un análisis en cuanto al nivel de profundidad, se trata de una investigación **exploratorio-descriptiva** puesto que, en primer lugar, indaga en conceptos generales para poder luego aproximarse a los objetivos finales de la investigación. En este último punto es cuando pasa a ser considerada **descriptiva** ya que, justamente, busca describir cómo se relacionan diversas variables asociadas al ordeño y la lechería.

Además, según la dimensión temporal, el presente trabajo es ***diacrónico*** porque trata la información teniendo en cuenta y analizando su evolución a lo largo del tiempo.

Haciendo referencia al tipo de dato con el que trabaja, esta investigación es del tipo ***de campo*** puesto que la información es tomada de una base de datos y luego, se realizan una serie de “experimentos”. Adicionalmente, también es considerada del tipo ***documental*** ya que generará un documento con valiosas conclusiones obtenidas.

De acuerdo al método empleado es considerada ***cuantitativa*** ya que se centra en obtener gráficas, curvas matemáticas, indicadores y diversos factores estadísticos y numéricos.

Adicionalmente se puede afirmar que se trata de una investigación ***pura***, ya que su finalidad es contribuir con el incremento de los conocimientos sobre la temática.

## Técnicas de recolección de datos

Gracias al convenio establecido entre UCSE-DAR y el INTA es que, éste último, accedió a proveer para el presente trabajo una réplica de la base de datos con la que opera el sistema del tambo robotizado en cuestión, en formato de backup de SQL Server.

En dicha réplica de la base de datos, se cuenta con un registro histórico de los ordeños diarios de 71 vacas. El rango de tiempo de estos datos abarca un período de 1 año aproximadamente, yendo desde el 14/03/2016 al 19/04/2017.

## Plan de tratamiento de datos

En primer lugar, se hará un análisis de toda la información con la que se cuenta en la base de datos. Luego de este proceso, se generarán diversos archivos *CSV* (del inglés *comma separated values*) con la información relevante para la investigación. Estos archivos son un tipo de documento en formato abierto sencillo para representar datos en forma de tabla, en las que las columnas se separan por comas y las filas por saltos de línea.

La idea de trabajar con este formato está fundamentada en la posibilidad de utilizarlos en cualquier tipo de programa con relativa sencillez, logrando así la independencia de la plataforma que se vaya a utilizar.

Una vez obtenidos estos archivos se generará un documento de *Jupyter Notebook*. Este tipo de archivos es conocido como un entorno computacional interactivo, en el cual es posible combinar texto enriquecido, ejecución de código Python, cálculos matemáticos y dibujo de gráficas representativas.

Para el procesamiento, filtrado y agrupado de la información se utilizará el lenguaje de programación *Python*, en combinación con *Pandas*.

Pandas es una librería de código abierto para el lenguaje de programación Python, con licencia BSD, que proporciona estructuras de datos de alto rendimiento, facilidad de uso y herramientas para el análisis de datos.

Con todas estas tecnologías mencionadas, entonces, será posible realizar el presente trabajo de investigación. Básicamente, se importarán los archivos CSV al Jupyter Notebook, se realizará todo el procesamiento y análisis necesario mediante código Python y, por último, se presentarán las gráficas obtenidas como resultado de la investigación.

# Alcance

En primer lugar, se establece como objetivo el análisis, filtrado, preprocesamiento y agrupación de la información para obtener así la curva de producción (en litros de leche) por intervalos (diferencia de tiempo entre ordeños sucesivos).

En segundo lugar, se pretende agregar filtros adicionales que permitan observar esta misma curva, pero en función de ciertos parámetros que pueden resultar de interés tales como, número de lactación de la muestra, tercio de lactancia, origen de los animales, cantidad de individuos aceptados por intervalos, etc.

En tercer lugar, se tratará de encontrar algún tipo de correlación entre los datos de manera que sea posible responder, de manera aproximada, a los interrogantes referidos a cuestiones tales como: cantidad de ordeños por día para la consecución de una máxima eficiencia, perdida de eficiencia debida a ordeños en intervalos irregulares, etc.

Por último, se plantea la posibilidad del desarrollo de un modelo de inteligencia artificial que, en función de la información disponible, logre aprender y predecir la producción esperada por animal en el próximo ordeño. Actualmente, el sistema cuenta con dicha predicción, pero se considera que tal vez pueda mejorarse el porcentaje de exactitud de la misma.

Es importante mencionar que, este último punto, tal vez pueda dejarse simplemente planteado para una futura investigación o continuación del presente trabajo.

# Cronograma

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Actividades-Tiempo | 1º mes | 2º mes | 3º mes | 4ºmes | 5ºmes |
| Importar la base de datos. |  |  |  |  |  |
| Analizar la información y generar los archivos CSV. |  |  |  |  |  |
| Eliminar datos con errores. |  |  |  |  |  |
| Obtener la curva de producción por intervalos. |  |  |  |  |  |
| Agregar filtros adicionales al documento interactivo. |  |  |  |  |  |
| Obtener, para cada registro, el penúltimo intervalo. |  |  |  |  |  |
| Obtener una gráfica representativa de la correlación entre un intervalo y el anterior. |  |  |  |  |  |
| Hacer un análisis de rendimiento en función de la cantidad de ordeños por día. |  |  |  |  |  |
| Aproximar un porcentaje de pérdida de eficiencia en función de la diferencia horaria entre ordeños. |  |  |  |  |  |

# Bibliografía

Bretschneider, G., Salado, E., Cuatrin, A., & Arias, D. (s.f.). *Lactancia: Pico y Persistencia.* Recuperado el 8 de Agosto de 2017, de http://inta.gob.ar/sites/default/files/script-tmp-inta\_lactancia\_pico\_y\_persistencia\_febrero\_2015.pdf

*Presentamos el tambo robotizado*. (18 de Noviembre de 2015). Recuperado el 26 de Agosto de 2017, de http://inta.gob.ar/noticias/presentamos-el-tambo-robotizado

Sampieri, R. H., Collado, C. F., & Lucio, P. B. (2006). Metodología de la Investigación. *Cuarta*. México: Mc. Graw Hill.