**PLAN DE ESTUDIOS CARRERA DE INGENIERIA DE TECNOLOGÍAS DE INFORMACIÓN Y SISTEMAS - 2020**

**Secuencia de Cursos para el área de Ciencia de Datos**

|  |  |  |
| --- | --- | --- |
| **CICLO** | **CURSO** | **OBSERVACIONES** |
| TERCER | ESTADÍSTICA Y PROBABILIDADES | El curso estaba programado en el segundo ciclo dejando un ciclo sin cursos de Estadística |
| CUARTO | ESTADÍSTICA INFERENCIAL |  |
| CUARTO | ALGEBRA LINEAL I | El curso estaba programado en el quinto ciclo, pero debe ser prerrequisito del curso de Análisis Multivariado I |
| QUINTO | ANÁLISIS MULTIVARIADO I | El curso antes se denominaba ANÁLISIS MULTIVARIADO |
| SEXTO | ANÁLISIS MULTIVARIADO II | El curso antes se denominaba PROGRAMACIÓN PARA ANÁLISIS ESTADÍSTICO |
| SEPTIMO | MINERÍA DE DATOS |  |
| OCTAVO | SERIES DE TIEMPO | Programar este curso en reemplazo del curso de Minería de Datos Avanzado |

Curso: ANÁLISIS MULTIVARIADO I

Créditos: 4 Requisito: Estadística Inferencial

Sumilla:

La asignatura de Análisis Multivariado, presenta las herramientas estadísticas fundamentales para abordar situaciones que involucran fenómenos gerenciales y de situaciones empresariales que involucran conjuntos de datos con múltiples variables, integra además los métodos descriptivos multivariantes, de reconocimientos de patrones, reducción de la dimensión y el análisis exploratorio de datos.

**El curso Análisis Multivariado I presenta los métodos básicos del análisis multivariado que permiten abordar situaciones relacionadas con fenómenos gerenciales y de situaciones empresariales usando conjuntos de datos de grandes dimensiones. Incluye métodos descriptivos multivariantes, técnicas de reducción de dimensiones, algoritmos para la identificación de grupos y construcción de modelos para explicar el comportamiento de una variable cuantitativa.**

El contenido de la asignatura comprende: Análisis de Componentes Principales, Análisis Factorial, Análisis de Correspondencias Simples, Análisis Cluster Jerárquico y de Partición, Regresión Múltiple, Análisis Discriminante Lineal y Análisis Discriminante Cuadrático.

Curso: ANÁLISIS MULTIVARIADO II

Créditos: 3 Requisito: Análisis Multivariado I, Programación Orientada a Objetos

Sumilla:

El curso permitirá al alumno determinar, analizar y describir las estructuras de datos estadísticos y su respectivo resumen numérico y gráfico a través de una herramienta de programación estadístico (Ejemplo R). Permitirá adquirir los conceptos, patrones de razonamiento y estrategias básicas de la inferencia estadística. Modelizar estadísticamente problemas reales, centrando el interés en el ajuste y validación del modelo, así como en su enfoque computacional a través del lenguaje de programación especializado. Definir estrategias básicas para atender problemas de reducción de dimensión de conjuntos de datos y selección de variables relevantes. Exponer métodos de búsqueda de agrupaciones o patrones en grandes masas de datos y técnicas de regresión.

**El curso Análisis Multivariado II presenta los métodos avanzados del análisis multivariado que permiten construir modelos para explicar el comportamiento de variables de naturaleza cuantitativa y cualitativa. Incluye las técnicas de remuestreo que son necesarias para una correcta evaluación de los modelos obtenidos y para el cálculo de los indicadores correspondientes.**

El contenido de la asignatura comprende: Regresión Logística Nominal, Indicadores para evaluar modelos de clasificación binaria, Regresión Logística Ordinal, Regresión Poisson, Métodos de Remuestreo, Regresión Ridge, Regresión Lasso, Regresión Elastic Net y Árboles de Regresión

Curso: MINERÍA DE DATOS

Créditos 3 Requisito: Inteligencia de Negocios, Análisis Multivariado II

Sumilla:

El curso examina a través de métodos que han emergido de disciplinas como la estadística y la inteligencia artificial y han probado ser muy efectivos al momento de reconocer patrones de comportamiento y determinación y reglas de inferencia.

Se basa en las etapas de descripción de datos, clasificación de datos, estimación de datos, predicción de datos y afinidad (agrupación) de datos.

Se harán uso de técnicas basadas en los principios de árboles de decisión, regresión logística, análisis discriminante y redes neuronales, entre otras.

Se apelará a herramientas especializadas en minería de datos.

Se deberá proponer proyectos integradores con inteligencia de negocios a los diferentes escenarios de la cadena de valor de las organizaciones.

El contenido de la asignatura comprende: Metodología de minería de dato, técnicas de preprocesamiento de datos, Tratamiento de datos desbalanceados, Selección de variables, K-NN, Naive Bayes, Máquinas de Soporte Vectorial, Arboles de clasificación con C5 y CART, Bagging, Random Forest, Boosting, Ensamble y Stacking.

Curso: SERIES DE TIEMPO

Créditos 3 Requisito: Minería de Datos, Inteligencia Artificial aplicada los negocios

Sumilla:

El contenido de la asignatura comprende: Componentes de una serie de tiempo, Promedios Móviles, Métodos de Suavización, Metodología Box-Jenkins (ARIMA). Análisis Espectral y Series de Tiempo con Redes Neuronales.

**Otras sugerencias a la malla 2020**

* Revisar los contenidos de los cursos de Inteligencia Artificial Aplicado a los Negocios y de Inteligencia Avanzada para que no haya duplicidad de temas que están siendo desarrollados en Análisis Multivariado I, Análisis Multivariado II y Minería de Datos.
* Hay temas de Aprendizaje No Supervisado como Reglas de Asociación y Sistemas de Reconocimiento de Patrones que deberían incluirse en los cursos de Inteligencia Artificial
* Hay temas de Aprendizaje No Supervisado como Cluster PAM, Cluster CLARA, Cluster por Densidades que deberían incluirse en los cursos de Inteligencia Artificial.
* El curso de Series de Tiempo propuesto me parece que es muy específicio para el perfil de la carrera. Tal vez se podría considerar como curso electivo, o los temas fundamentales incluirlos en algún otro curso donde se enseñan técnicas predictivas.
* Los temas de análisis de datos no estructurados (e.g. minería de textos) no se están considerando en ninguno de los cursos de Minería de Datos ni de Inteligencia Artificial. Se debería corroborar con el syllabus del curso de Análitica en la Web. En caso no se estén tocando, sería una buena alternativa para considerarlos en el curso de Minería de Datos Avanzado o Minería de Datos II.
* La propuesta de contenido del curso de Minería de Datos siento que es un poco amplia. Quizás se debería ahondar más en la parte teórica de los diferentes modelos, ya que ahora los alumnos tendrán mejores fundamentos matemáticos con los cursos de Álgebra Lineal que se incorporarán. Esto permitirá que conozcan lo que hay detrás de cada algoritmo y les permitirá tener mejor dominio de la técnica. Tal vez el curso de manera más minuciosa podría llegar hasta el modelo Random Forest, y los modelos restantes podrían llevarse en un siguiente curso de Minería de Datos o en algún otro curso de Inteligencia Artificial, tomando en cuenta que Minería de Datos sea un requisito para el curso de Inteligencia Artificial.