 <b>UNIVERSIDAD SANTO TOMÁS</b> PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA	<b>PLAN DE TRABAJO DE LA ASIGNATURA</b>		
Código:	Versión: 01	Emisión:	Página 1 de 5

**PERÍODO**  **ASIGNATURA**

**INTENSIDAD:** **Clase:**  **Independiente:**  **Total:**

**CÓDIGO:**  **CRÉDITOS:**

**HORARIO:**

**NOMBRE DEL DOCENTE:**

**CORREO ELECTRÓNICO:**

## 1. Justificación

Bajo distintos nombres según el dominio—inteligencia de negocios, analítica de datos, *big data*, inteligencia artificial, minería de datos, *machine learning*—, las técnicas de aprendizaje computacional basado en datos son una alternativa presente y poderosa de modelamiento predictivo de fenómenos complejos. Es necesario proveer al estudiante de la Maestría en Estadística aplicada de las herramientas conceptuales, algorítmicas y de programación que le permitan proponer y evaluar soluciones eficientes y precisas a problemas de aprendizaje automático sobre todo tipo de datos.


## 2. Objetivos

### 2.1. General

Introducir una taxonomía del tipo de problemas presentes en minería de datos y *machine learning* y familiarizar al estudiante con los enfoques, métodos, algoritmos y demás herramientas computacionales disponibles para resolverlos.

### 2.2. Específicos

1. Ayudar al estudiante a interiorizar un procedimiento global de solución de problemas con base en datos, que enfatice la precisión y eficiencia de las respuestas y las buenas prácticas metodológicas y de presentación de resultados.
2. Capacitar al estudiante en distintos métodos de aprendizaje supervisado, semi- y no supervisado para la solución de problemas de clasificación, regresión y similares basados en datos.
3. Proveer herramientas para la valoración del desempeño de algoritmos de aprendizaje en aplicaciones específicas.

 <b>UNIVERSIDAD SANTO TOMÁS</b> PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA	<b>PLAN DE TRABAJO DE LA ASIGNATURA</b>		
Código:	Versión: 01	Emisión:	Página 2 de 5

4. Suministrar al estudiante las bases matemáticas y conceptuales para entender e implementar algoritmos de solución diferentes a los cubiertos en el espacio académico.

### 3. Competencias académicas

El estudiante que apruebe este curso tendrá una ventaja competitiva bastante clara con respecto a otros de su misma área pues fortalecerá, entre otras, competencias como la lecto-escritura, mediante la realización de informes; la capacidad argumentativa, a través de la selección apropiada de la metodología de aprendizaje con base en los datos disponibles; interpretativa, a través de la comprensión de artículos con contenido estadístico, matemático, algorítmico y de implementación; y lógico-matemática, mediante el análisis de los resultados generados a partir de la aplicación de técnicas de minería de datos y *machine learning* y su aplicación práctica trans-disciplinar, además de las aplicaciones con bases de datos reales en diferentes paquetes estadísticos.

### 4. Metodología

Los aspectos teóricos se desarrollarán en clases magistrales y en paralelo se realizan las prácticas, en sesiones de trabajo con el computador. De manera transversal al curso, el estudiante desarrollará competencias de programación computacional, debido a que todas las prácticas se realizarán en el software estadístico **R**.

### 5. Evaluación

#### 5.1. Asistencia


La asistencia y puntualidad en clase son determinantes a la hora de la calificación final. El estudiante que reporte fallas en más del 20 % del curso (tiempo) perderá automáticamente.

#### 5.2. Actividades

Durante el transcurso del curso, el estudiante tendrá que presentar un examen parcial, una prueba práctica final, talleres prácticos en clase y tareas conceptuales, de derivación matemática y algorítmica y, excepcionalmente, de programación.

La nota final del curso estará determinada así:


- 20 % – Talleres prácticos en clase – Aproximadamente uno por sesión.
- 20 % – Tareas de análisis/derivación/programación – Aproximadamente una por semana.
- 20 % – Examen parcial – Viernes 4 de mayo, desde las 10:00 p.m.
- 40 % – Prueba final – Sábado 19 de mayo desde las 7:00 a.m.

 <b>UNIVERSIDAD SANTO TOMÁS</b> PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA	<b>PLAN DE TRABAJO DE LA ASIGNATURA</b>		
Código:	Versión: 01	Emisión:	Página 3 de 5

## 6. Contenido programático

El desarrollo semanal propuesto para el curso es el siguiente:

Sem.	Temas y subtemas	Trabajo presencial	Trabajo de acompañamiento	Trabajo independiente	Recursos y/o material de apoyo
1	Introducción: Taxonomía de problemas sobre datos y enfoques de solución desde analítica y aprendizaje computacional. Elementos de optimización y álgebra lineal. Aspectos teóricos generales.  Métodos básicos de regresión y clasificación.	Clase magistral y construcción conjunta de un ejemplo exhaustivo.	Desarrollo práctico en sala.	Preguntas y ejercicios de práctica.	Ejercicios suministrados por el docente.
2	Derivación y aplicaciones de análisis de componentes principales desde la perspectiva de datos: reducción de dimensionalidad, análisis de anomalías, proyección al espacio nulo, rellenado de datos.  Métodos de <i>clustering</i> : <i>k-means</i> , jerárquico, <i>SOM</i> .	Clase magistral, herramientas para aplicación de los métodos a diferentes escalas.	Desarrollo práctico con datos de imágenes y generados artificialmente.	Preguntas y ejercicios de práctica.	Ejercicios suministrados por el docente.
3	Tipología de problemas de aprendizaje supervisado: clasificación binaria y multi-etiqueta, análisis sensible al costo, aprendizaje en línea.  <i>Support Vector Machines</i> : derivación, interpretación y ejemplos.	Clase magistral y práctica en sala.	Desarrollo práctico: Clasificación multi-etiqueta sensible al costo en datos artificiales. Clasificación datos MNIST usando <i>SVMs</i> .	Preguntas y ejercicios de práctica.	Ejercicios suministrados por el docente.

 <b>UNIVERSIDAD SANTO TOMÁS</b> PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA	<b>PLAN DE TRABAJO DE LA ASIGNATURA</b>		
Código:	Versión: 01	Emisión:	Página 4 de 5

4	Redes neuronales: algoritmo de <i>backpropagation</i> , algunos resultados teóricos y aplicaciones.  Fundamentos de <i>Deep Learning</i> : <i>autoencoders</i> , redes convolucionales, herramientas de implementación a mediana escala.	Clase magistral y práctica en sala.	Desarrollo práctico sobre datos MNIST y otros.	Preguntas y ejercicios de práctica.	Ejercicios suministrados por el docente.
5	Técnicas de combinación de modelos: bosques, <i>boosting</i> , comités.  Prueba exhaustiva final.	Clase práctica y magistral.  Desarrollo práctico.	Desarrollo práctico en sala de sistemas.	Preguntas y ejercicios de práctica.	Ejercicios suministrados por el docente.  Prueba Final.


## 7. Bibliografía

### 7.1. Texto guía

- Bishop, C.M., (2006), Pattern Recognition and Machine Learning. Primera Edición. Springer, Estados Unidos.

### 7.2. Bibliografía básica

- James, G., Witten, D., Hastie, T. y Tibshirani, R., (2013), An Introduction to Statistical Learning with Applications in R. Primera Edición. Springer, Estados Unidos.
- Witten, I.H., Frank, E. y Hall, M.A., (2011), Data Mining: Practical Machine Learning Tools and Techniques. Tercera Edición. Morgan Kaufman, Estados Unidos.
- Schölkopf, B. y Smola, A.J., (2002), Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Primera Edición. The MIT Press, Estados Unidos.
- Haykin, S., (2009), Neural Networks and Learning Machines. Tercera Edición. Pearson Prentice Hall, Estados Unidos.

 <b>UNIVERSIDAD SANTO TOMÁS</b> PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA	<b>PLAN DE TRABAJO DE LA ASIGNATURA</b>		
Código:	Versión: 01	Emisión:	Página 5 de 5

### 7.3. Bibliografía complementaria

- Kuncheva, L.I., (2004), Combining Pattern Classifiers: Methods and Algorithms. Primera Edición. Wiley-Interscience, Estados Unidos.
- Kirby, M., (2000), Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns. Primera Edición. Wiley-Interscience, Estados Unidos.