



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



**APRUEBA LA CARRERA DE MINERÍA DE DATOS EN EL ÁMBITO DE LA
UNIVERSIDAD TECNOLÓGICA NACIONAL**

Buenos Aires, 27 de abril de 2017

VISTO la decisión de jerarquizar la educación de posgrado en la Universidad Tecnológica Nacional, abarcando los diferentes niveles y aspirando al mayor reconocimiento nacional e internacional, y

CONSIDERANDO:

Que el campo temático que aborda la carrera de Maestría en Minería de Datos resulta un área de conocimiento que es imprescindible ofrecer a través de la formación de posgrado.

Que la sociedad del siglo XXI es productora de una enorme cantidad de información que para ponerla en valor es necesario clasificar, jerarquizar y modelizar.

Que la Universidad pretende aportar desde lo académico, científico y profesional al desarrollo de los tópicos que abordan el análisis avanzado de datos a través de la estadística, la inteligencia artificial y las bases de datos.

Que las temáticas de la carrera Minería de Datos reviste importancia en áreas tales como estrategias de marketing, soporte de decisiones, planeamiento financiero y análisis de datos científicos en ciencias biomédicas, sociales, económicas y comunicaciones, entre otras.

Que la formación de profesionales en el campo de la Maestría se enmarca en el área de informática, un área de fuerte crecimiento en los últimos años y gran demanda insatisfecha en el país.



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



Que se espera que la Maestría en Minería de Datos desarrolle competencias tendientes a expandir el campo científico y profesional, producir bienes científicos y tecnológicos y fortalecer el desarrollo económico y social.

Que la Comisión de Posgrado de la Universidad avala la propuesta y la Comisión de Ciencia, Tecnología y Posgrado recomiendan su aprobación.

Que el dictado de la medida se efectúa en uso de las atribuciones otorgadas por el Estatuto Universitario.

Por ello,

EL CONSEJO SUPERIOR DE LA UNIVERSIDAD TECNOLÓGICA NACIONAL

ORDENA:

ARTICULO 1°.- Aprobar la carrera de Maestría en Minería de Datos que se agrega como Anexo I y es parte integrante de la presente ordenanza.

ARTÍCULO 2°.- Aprobar el diseño curricular de la mencionada carrera, que se agrega como Anexo I y es parte de la presente Ordenanza.

ARTÍCULO 3°.- Dejar establecido que su implementación en la Universidad, a través de sus Facultades Regionales, debe ser expresamente autorizada por el Consejo Superior cuando se cumplan las condiciones y los requisitos estipulados en las normativas que rigen la educación de posgrado de la Universidad Tecnológica Nacional.

ARTÍCULO 4°.- Regístrese, comuníquese y archívese.

ORDENANZA N° 1584

UTN
SCTYP
l.p.
f.c.r.

A.U.S. RICARDO F. O. SALLER
Secretario del Consejo Superior

Ing. HÉCTOR CARLOS BROTO
RECTOR



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



ORDENANZA N° 1584

ANEXO I

MAESTRÍA EN MINERÍA DE DATOS

I. FUNDAMENTACIÓN

En los últimos años se han acumulado enormes cantidades de datos en todas las organizaciones y esta tendencia continúa a un ritmo acelerado. Esto ha sido posible dado el amplio uso de sistemas computarizados, nuevas técnicas de captura de datos, el empleo de códigos de barra, los lectores de caracteres ópticos, las tarjetas magnéticas, etc. y por el avance en la tecnología de almacenamiento y su consiguiente reducción de costos. El sentido de conservar todos esos datos surge del convencimiento de que los mismos representan un importante activo para cualquier organización que los posea, en la medida en que puedan ser transformados en información útil.

Se estima que, del análisis de esos datos, pueden surgir ventajas competitivas o novedosas soluciones a antiguos problemas.

La Minería de Datos (Data Mining), también referenciado como Descubrimiento del Conocimiento en Bases de Datos (Knowledge Discovery in Databases o KDD), ha sido definida como el proceso de extracción no trivial de información implícita, previamente desconocida y potencialmente útil.

El crecimiento explosivo de las bases de datos y el aún mayor de la Internet, nos urge en la búsqueda de técnicas y herramientas que, de manera automática y eficiente, generen información, a partir de los datos almacenados. Este es el objetivo de las técnicas de Data Mining y Knowledge Discovery.

En particular, estas técnicas han adquirido enorme importancia en áreas tales como



*Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado*



estrategias de marketing, soporte de decisiones, planeamiento financiero y el análisis de datos científicos, entre otras.

El origen de la Minería de Datos se puede encontrar, básicamente, en la intersección de áreas del conocimiento tales como la Estadística, la Inteligencia Artificial (Machine Learning) y Bases de Datos.

Asimismo, gran cantidad de organizaciones, tanto públicas como privadas, con interés comercial o científico, han acumulado enormes cantidades de datos, estimando lograr ventajas competitivas o novedosas soluciones como resultado de la explotación de los mismos.

En resumen, Minería de Datos es un conjunto de técnicas de extracción de modelos, patrones, recurrencias, anomalías y otras estructuras de significación estadística desde grandes bases de datos y cuyo interés radica en el hecho que los datos se han acumulado en forma exponencial mientras que el número de analistas lo ha hecho en forma lineal.

II. JUSTIFICACIÓN

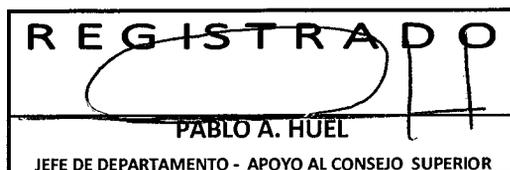
La carrera se enmarca en el área de informática, en un área de fuerte crecimiento en los últimos años y gran demanda insatisfecha en el país, constituyendo esta propuesta la primera fuera de Capital Federal.

La carrera está destinada a egresados de diversas carreras de Ingeniería y de otras carreras del país que requieran formación específica en esta temática: Contador Público, Economía, Administración, Actuarios, Medicina, Biología, Bioingeniería, Bioinformática, Agronomía, Climatología, por citar sólo algunas.

La Maestría está dirigida a profesionales de organizaciones en sectores como los de los servicios, finanzas, telecomunicaciones, gobierno, manufactura, industria alimenticia, etc., responsables de gestionar el conocimiento del negocio basado en la explotación de datos, en cargos tales como ingenieros asesores a empresas, analistas de inteligencia de



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



mercado, responsables de áreas de inteligencia en gestión de negocios, jefes de investigación de mercado, jefes de gestión y análisis de bases de datos, gerentes de información de negocios, responsables de investigación comercial, responsables de áreas de marketing apoyado en bases de datos, directores de negocios, gerentes y personal de empresas aseguradoras entre otras actividades similares.

Otras áreas de interés son las dedicadas específicamente a investigación científica y tecnológica: sistemas de geoposicionamiento satelital, análisis de datos de siembra, cosecha y otras actividades agropecuarias, análisis de situaciones climáticas, monitoreo de diversas situaciones sociales: mapas de violencia, de seguridad, monitoreo de desastres naturales, manejo de pacientes en hospitales, seguimiento de enfermedades, tráfico de comunicaciones y volumen de intercambio de datos telefónicos y por internet, movimiento de aves y peces, por citar sólo unas pocas actividades.

Muchas empresas e instituciones requieren de técnicas de análisis de minería de datos para mejorar su rendimiento y productividad sin saber que disponen de estas metodologías.

III. OBJETIVOS DE LA CARRERA

Formar recursos humanos altamente capacitados en los fundamentos y el conocimiento práctico, que permitan tanto la aplicación, de manera creativa y rigurosa de los métodos de Minería de Datos y Descubrimiento del Conocimiento en un marco interdisciplinario y con conceptos e instrumentos avanzados, como la producción de conocimientos científicamente validados en el ámbito de dicha disciplina.

- Generar y mantener actividades de investigación, desarrollo y transferencia tecnológica en el área de explotación y análisis de datos.
- Contribuir a la generación y actualización permanente de recursos humanos para el ámbito universitario en estas temáticas.



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



- Conceptualizar modelos representativos sobre situaciones de todo tipo.
- Comprender el ciclo de trabajo de las Ciencias de Datos y de la Gestión del Conocimiento dentro de las organizaciones y saber gestionarlo con éxito.

IV. PERFIL DEL GRADUADO

Se espera que quien se gradúe de esta Maestría adquiera competencias que le permitan:

- Manipular, explorar, limpiar y preparar cualquier fuente de datos antes de su procesamiento.
- Diseñar, desarrollar e implementar técnicas de Aprendizaje de Máquinas para su utilización aplicada a través de modelos predictivos, sistemas de recomendación, scoring, reconocimiento de segmentos y clusters, entre otras.
- Diseñar visualizaciones de información.
- Comunicar eficazmente los hallazgos traduciéndolos de manera comprensible a personas no especializadas.
- Construir secuencias de análisis con R y Python, usando poderosas librerías de análisis y eventualmente construir sus propias librerías.
- Utilizar las herramientas de trabajo diario del profesional de minería de datos.
- Participar activamente de las comunidades científicas y trabajar en el desarrollo colaborativo en proyectos de innovación.
- Desempeñarse con éxito como analista científico de datos en cualquier organización, empresa o emprendimiento en cualquier rubro.

V. NOMBRE DE LA CARRERA Y TÍTULO QUE OTORGA

La carrera se denomina "Maestría en Minería de Datos" y el título que otorga es el de "Magíster en Minería de Datos".



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



VI. NORMAS DE FUNCIONAMIENTO

Condiciones de ingreso

Podrán ser inscriptos en la Maestría en Minería de Datos aquellos profesionales que posean título superior de grado de ingeniero, licenciado o equivalente, otorgado por Universidad reconocida.

En todos los casos se realizará una evaluación de los candidatos al ingresar al programa para determinar el grado de correspondencia entre su formación, trayectoria y los requisitos de la carrera. La evaluación se realizará a través del análisis de antecedentes, entrevistas, y eventualmente, de otros elementos de juicio solicitados por el Director de la carrera y Comité Académico. En caso de ser necesario se realizará un coloquio a cargo de los mismos. También podrán recomendarse cursos de nivelación para los aspirantes que requieran profundizar o actualizar sus conocimientos en áreas específicas.

Condiciones de admisión

La admisión como maestrando está a cargo del Consejo Superior o del Consejo Directivo según corresponda. La Comisión de Posgrado de la Universidad o de la Facultad Regional, evaluará los siguientes componentes:

- a. Plan de Trabajo de tesis avalado por el director de tesis propuesto;
- b. Curriculum vitae del director y codirector de tesis (si corresponde);
- c. Curriculum vitae del tesista en el que se detalle, si las hubiera, las tareas de investigación y desarrollo, publicaciones, cursos y seminarios de posgrado así como otros antecedentes referidos a la temática central de la tesis propuesta.

En todos los casos se requerirá el manejo del idioma inglés.

Modalidad

El régimen de cursado previsto es presencial y se deben cumplimentar los contenidos y las cargas horarias mínimas establecidas para los cursos y seminarios que integran el plan de



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



estudios. En el caso de utilización de metodologías de educación a distancia, su uso no deberá exceder el 30% del total de horas presenciales de la carrera.

Evaluación y Promoción

La promoción supone asistencia regular a las clases – con una asistencia mínima del ochenta por ciento (80%) -, presentación adecuada de trabajos y/o tareas solicitadas por los responsables académicos de los cursos y aprobación de las evaluaciones previstas.

La evaluación ligada a la promoción y acreditación o sumativa informa sobre los logros alcanzados por los alumnos y califica su rendimiento en términos de los objetivos alcanzados por ellos por medio de diferentes instrumentos: pruebas parciales, coloquios integradores, informes, monografías, exámenes finales individuales y presenciales.

La calificación se expresará en escala numérica de CERO (0) a DIEZ (10) sin decimales. Para la promoción se requerirá la nota mínima de SIETE (7) en evaluaciones individuales y presenciales.

Metodología de enseñanza

La formación estará centrada en la articulación entre los conocimientos propios del campo de estudio, la experiencia profesional previa y la aplicación de los conocimientos adquiridos en casos concretos. La propuesta de enseñanza se estructura en torno a las siguientes estrategias:

- Desarrollo de los temas a través de exposición, diálogo, discusión y uso de variadas metodologías colaborativas que promuevan la apropiación del conocimiento.
- Experiencias de práctica experimental en laboratorios.
- Utilización de estrategias como: estudio de casos, proyectos, simulación, resolución de problemas, visitas y trabajos de campo.
- Talleres y seminarios.

Los profesores responsables del dictado de los cursos y seminarios podrán solicitar la



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



presencia de otros profesores en carácter de invitados, con similares antecedentes académicos y profesionales, para el desarrollo de las temáticas teóricas, la comunicación de investigaciones y la presentación de planteos metodológicos y técnicos vinculados con los contenidos particulares a tratar.

Requisitos de graduación

Para acceder al Título de Magíster en Minería de Datos, el maestrando deberá cumplir los siguientes requisitos:

- Aprobar una prueba de suficiencia de idioma inglés.
- Cumplir con la carga horaria mínima de horas establecidos para la carrera.
- Culminar los estudios en el tiempo máximo fijado por el Reglamento de Posgrado vigente en la Universidad.
- Aprobar la defensa de la tesis.

La tesis consistirá en un trabajo de investigación o en un desarrollo o aplicación de conceptos o procedimientos a un problema dado, que demuestre destreza en el manejo conceptual y metodológico acorde con el estado actual del conocimiento y un tratamiento innovador, ya sea en la metodología o en los enfoques teóricos vinculados a la temática de la carrera.

Como lo establece el Reglamento de la Educación de Posgrado de la Universidad, es posible solicitar reconocimiento de créditos académicos a través de otros cursos, seminarios y/o carreras de posgrados afines. La Comisión de Posgrado analizará la pertinencia, la relevancia y la profundidad de los contenidos y las características de la entidad u organismo donde se desarrollaron los tópicos antes indicados, estableciendo la cantidad de créditos académicos a reconocer.

Duración

El plazo máximo para cumplir con todas las obligaciones del Plan de Estudios incluida la defensa de la tesis, es de CUATRO (4) años. En la eventualidad que este período sea



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



vencido, y ante solicitud fundamentada, el Consejo Superior o el Consejo Directivo de la Facultad Regional, según corresponda, podrá conceder una prórroga de manera excepcional, que no podrá ser superior a UN (1) año.

Financiamiento

La carrera deberá autofinanciarse y se desarrollará a través de las Facultades Regionales que lo soliciten; las cuales se harán responsables por la inscripción, la recepción de solicitudes, el cobro de aranceles y fijación del monto de los mismos, así como del apoyo técnico administrativo.

Organización Académica

Las Facultades Regionales autorizadas por el Consejo Superior a poner en vigencia y ofrecer la "Maestría en Minería de Datos" deberán establecer las figuras institucionales de Dirección de la Carrera y Comité Académico que serán responsables de:

- Establecer los lineamientos y las orientaciones para el desarrollo curricular de la carrera.
- Seleccionar y proponer a los integrantes del Cuerpo Docente.
- Evaluar los programas analíticos de los cursos y seminarios.
- Evaluar el desempeño de docentes y estudiantes.
- Efectuar el seguimiento académico de la implementación de la carrera.
- Participar en las entrevistas y evaluar las condiciones de los aspirantes para su admisión.
- Orientar a los estudiantes en la elección de los temas de tesis y en el desarrollo de las actividades conducentes a la finalización de la tesis.
- Entender en el proceso de revisión y actualización de la carrera.

VII. ESTRUCTURA CURRICULAR Y SU ORGANIZACIÓN

La carrera consta de catorce (14) cursos que cubren un total de 648 horas de cursado. Doce (12) de los cursos son de carácter obligatorio y dos (2) de carácter optativo, a elección entre



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



un menú de cinco (5) que se ofrecen, sin perjuicio que eventualmente se pueda incorporar otras asignaturas optativas, previo a cumplir con los trámites de aprobación por los órganos pertinentes. Las horas reloj que corresponden a cada espacio curricular son teórico-prácticas, tal como lo establece la normativa vigente. Las actividades prácticas cubren entre el 25 y el 30 % del total y podrán variar de cátedra en cátedra. Dichas actividades deben consignarse en los programas analíticos de cada curso y pueden adoptar la modalidad de talleres, trabajos de campo, simulaciones, pasantías, entre otros. Se podrán incorporar otras actividades de formación, no comprendidas en el plan de estudio.

La carrera se estructura en cuatro bloques temáticos:

- Básicas, para dar bases de nivelación a los alumnos/as. Estos tres cursos podrán rendirse en forma libre, para aquellas personas que posean conocimientos sólidos en estas temáticas.
- Operativas, para acompañar el proceso formativo, consistentes talleres de plan de tesis/metodología de investigación y redacción científica.
- Fundamentos, donde se brinda el núcleo de conocimiento disciplinares.
- Optativas o de profundización, para que cada estudiante complete su formación de acuerdo a sus intereses.
- Finalmente, los estudiantes deberán acreditar 160 hs. destinadas a actividades de elaboración de la tesis y complementarias.



Ministerio de Educación y Deportes
 Universidad Tecnológica Nacional
 Rectorado



Plan de estudios

Bloque	Cursos/Talleres	Horas	SubTotal
Básicas	Algoritmos	24	
	Estadística	24	
	Base de Datos	24	
			72
Operativas	Taller de Tesis I	48	
	Taller de Tesis II	48	
			96
Fundamentos	Minería de Datos	64	
	Aprendizaje Automático	64	
	Análisis Inteligente de Datos	64	
	Aplicaciones de Minería de Datos a Economía y Finanzas	48	
	Aplicaciones de Minería de Datos a Ciencia y Tecnología	48	
	Enfoque Estadístico del Aprendizaje	48	
	Visualización de la Información	48	
			384
Optativas	Redes Neuronales	48	
	Minería de Texto	48	
	Datwarehouse	48	
	Sistemas de Información Geográfica	48	
	Series de Tiempo	48	
			96
Total de horas a acreditar de cursos presenciales			648
Trabajo de Tesis y Actividades complementarias			160
CARGA HORARIA TOTAL			808



Ministerio de Educación y Deportes
 Universidad Tecnológica Nacional
 Rectorado



Esquema de correlatividades previsto

<i>Para cursar</i>	<i>N°</i>	<i>Necesita tener aprobada</i>
Aprendizaje Automático		Algoritmos
Minería de Datos		Bases de Datos
Análisis Inteligente de Datos (AID)		Estadística
Aplicaciones Econ. y Finanzas		Minería de Datos y AID
Aplicaciones Ciencia y Técnica		Minería de Datos y AID
Enfoque Estadístico del Aprendizaje		Estadística

Se ofrecerán Talleres de Software que tendrán como propósito proporcionar apoyo a los estudiantes de la carrera que lo necesiten. Consisten en una serie de charlas y actividades prácticas en aula de informática, para el manejo de los softwares correspondientes. No son obligatorios, no otorgan créditos ni se computan como horas formales de la carrera sino que sólo sirven de soporte. Por ello no se proporcionan programas. Se propone una duración de 24 horas para cada Taller, pero esta carga horaria puede variar. Los Talleres de Software mencionados son: WEKA – SPSS – SAS – INFOSTAT – R – PHYTON.

OBJETIVOS Y CONTENIDOS MÍNIMOS DE LOS CURSOS Y TALLERES

o ALGORITMOS

Objetivos:

Se propone brindar conocimiento de programación y algunos algoritmos típicos para buscar, ordenar o clasificar elementos, así como representar datos vía un lenguaje de especificación o de programación.

Contenidos mínimos:

Conceptos básicos de programación. Tipos de datos: primitivos. El tipo de datos lista, arboles, grafos - Funciones sobre listas. Conceptos básicos de los programas imperativos.



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



Estructuras de control - variables. Funciones y procedimientos - pasaje de parámetros.
Conceptos de complejidad algorítmica. Recursión.

o *ESTADÍSTICA*

Objetivos:

Proveer los conocimientos básicos de probabilidad, distribuciones y métodos gráficos exploratorios para comprender y abordar los temas de Análisis Inteligente de datos y Enfoque Estadístico del aprendizaje

Contenidos mínimos:

Etapas de una investigación. Estadística Descriptiva. Definiciones: Población, muestra, Variable, Valor de una variable, Dato, Observación o Medición, Caso. Gráficos: Histograma: construcción y análisis de posibles formas. Medidas de Resumen: Medidas de Posición o Centrado: Promedio o Media Muestral, Mediana Muestral, Media α - Podada. Medidas de Dispersión o Variabilidad: Rango Muestral, Varianza Muestral. Desvío Estándar Muestral, Coeficiente de Variación, Distancia Intercuartil, Distancia entre Cuartos, Desvío Absoluto Mediano. Box-Plots. Construcción y análisis. Outliers. QQ-plot. Inferencia estadística: Método de momentos. Intervalos de Confianza. Construcción e interpretación. Intervalos de confianza para los parámetros de una distribución normal. Método general para obtener intervalos de confianza. Intervalos de confianza de nivel asintótico. Tests de hipótesis. Construcción e interpretación. P-valor. Error tipo I y II. Nivel de significación del test. Tests de hipótesis de nivel α para los parámetros de la distribución normal. Región de rechazo. Tests de hipótesis asintótico para la media de una distribución cualquiera. Relación entre tests de hipótesis bilaterales e intervalos de confianza.



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



o *BASES DE DATOS*

Objetivo:

Se espera que los alumnos puedan trabajar con diferentes tipos de datos y convertir entre varios formatos de bases de datos, importar y exportar tablas en diferentes formatos (Microsoft Excel, Fuentes de datos ODBC, Archivos separados por comas y tabuladores, Reconocer los caracteres especiales). Manejo básico de SQL. Manejo básico de consultas a bases de datos.

Contenidos mínimos:

Qué es una base de datos. Diferencia entre bases de datos y almacenamiento en archivos. Distintos tipos de DBMS (database management systems). Modelos de Datos. Modelos semánticos. Modelo Entidad/Relación. Diseño de una base de datos. Casos de estudio. Diferencias con UML. Modelos lógicos. Modelo relacional. Equivalencias entre modelos. Pasaje del modelo E/R al relacional. Lenguajes de Consulta. Álgebra Relacional. Operadores básicos: Selección, Proyección, Producto Cartesiano, Unión, Diferencia. Operadores complejos: Junta Natural, Cociente, Intersección. SQL como Lenguaje de Consulta. Transformación de AR a SQL. Cláusulas SELECT, FROM, WHERE, GROUP BY. Operadores avanzados. Diseño Relacional. Normalización. Tercera Forma Normal, Forma Normal de Óbice-Codd. Descomposición sin pérdida de información. Dependencias funcionales. Axiomas de Armstrong.

o *MINERÍA DE DATOS*

Objetivos:

La materia estará orientada a conseguir que el alumno sea capaz de identificar los problemas que se producen en el cruce de información al importar datos provenientes de distintas fuentes y tomar decisiones tendientes a mitigar los efectos de esos problemas.



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



Explicar el funcionamiento de los algoritmos de reglas de asociación y sequential patterns
Analizar las mejoras / variantes propuestas a los algoritmos de reglas de asociación / sequential patterns. Resolver problemas de data Mining aplicando reglas de asociación/sequential patterns

Contenidos mínimos:

Presentación General de todos los algoritmos de minería de datos. Proceso de Descubrimiento del Conocimiento. Tipos de datos. Mecanismos de limpieza de los mismos. Reglas de asociación. Patrones secuenciales. Aspectos temporales de las reglas de asociación. Otras aplicaciones de las reglas de asociación.

o *TALLER DE TESIS I*

Objetivos:

El taller está orientado a que las personas participantes reflexionen sobre su tema de tesis (o lo definan en caso de carecer de él), con revisión de aspectos de originalidad, factibilidad, coherencia, aplicabilidad, dimensión temporal, geográfica y de volumen del trabajo propuesto. Se trabajarán todos los pasos necesarios para la obtención de un proyecto o plan de tesis.

Contenidos mínimos:

Objeto de la tesis. Plan o proyecto de tesis. Criterios de evaluación del plan. Introducción, marco teórico, estado del arte. Factibilidad. Impacto. Objetivos. Hipótesis. Metodología. Equipo de dirección. Búsqueda bibliográfica. Conceptos de investigación científica.

o *APRENDIZAJE AUTOMÁTICO*

Objetivos:

El objetivo del dictado de esta materia es que el alumnado adquiera los conocimientos generales y prácticos del estado del arte en Aprendizaje Automático para su aplicación en la



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



práctica profesional en Minería de Datos. Por medio del curso el alumno deberá:

- Adquirir los conceptos básicos del aprendizaje computacional, sus fundamentos teóricos y una aproximación a sus bases formales;
- Conocer los principales modelos y algoritmos de aprendizaje computacional;
- Conocer metodologías que permitan seleccionar el modelo apropiado a los casos prácticos que se le presenten;
- Internalizar los conceptos por medio de la práctica con herramientas de libre disponibilidad establecidas en el mundo académico

Contenidos mínimos:

Introducción conceptual al aprendizaje automático. Introducción inductiva al procesamiento no tradicional de datos. Estructura de sistemas de aprendizaje y notación estándar. Métodos inferenciales no deductivos, generalización y noción de orden parcial. Algoritmos: Espacio de Versiones, FIND-S y Eliminación de Candidatos. Inducción de árboles de decisión. Algoritmo ID3 y derivados. Sesgo, ruido, sobre-especificación, sobre-generalización y poda. Aprendizaje como búsqueda heurística. Algoritmo STAR. Indicadores de performance y error. Aprendizaje basado en instancias y aprendizaje Bayesiano. Aprendizaje no supervisado. Clustering. Algoritmos aglomerativos y de partición. Introducción al Text Mining, aproximaciones NLP y ML. Combinación de múltiples modelos: Votación, Bagging, Boosting. Otros modelos de aprendizaje: SVM, Vere, Genéticos, Redes Neuronales, Analíticos, con teorías de dominio, basados en explicaciones, etc. Nuevas tendencias en Aprendizaje Automático.

○ *ANÁLISIS INTELIGENTE DE DATOS*

Objetivos:

Los Métodos del Análisis Exploratorio de Datos tienen por objetivo proveer una síntesis



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



global de un conjunto de datos sin hipótesis previas. El principio que guía este curso es introducir a los estudiantes en técnicas estadísticas multivariadas, y familiarizarlos con el software que se utiliza en sus aplicaciones y en la resolución de problemas diversos.

Contenidos mínimos:

Análisis exploratorio y confirmatorio. Reseña histórica. Revisión de métodos exploratorios; tablas, gráficos, diagramas de tallo y hoja, box-plot, análisis de normalidad. Técnicas descriptivas multidimensionales. Análisis en componentes principales. Análisis factorial de correspondencias. Métodos de clasificación y agrupamiento. Clasificación Jerárquica. K-medias. Caras de Chernov. Gráficos de estrellas. Gráficos de Rayos Sol. Gráficos de Andrews. Relación entre Análisis factorial y Clasificación. Árboles de decisión. Métodos de segmentación. Análisis discriminante. Análisis de datos.

○ *APLICACIONES DE MINERÍA DE DATOS A ECONOMÍA Y FINANZAS*

Objetivos:

Que los alumnos aprendan a desarrollar análisis de data mining que sean competitivos con los de nivel profesional existentes en el mercado argentino.

Contenidos mínimos:

Metodologías CRISP, Six sigma y SEMMA. Comparación de modelos. Nociones de Database Marketing. DataSets desbalanceados. Overfitting.

○ *APLICACIONES DE MINERÍA DE DATOS A CIENCIA Y TECNOLOGÍA*

Objetivos:

Este curso tiene como propósito presentar a los estudiantes los problemas y aplicaciones de minería de datos en algunas áreas de relevancia en ciencia y tecnología. Para lograr este objetivo es necesaria una breve introducción de los conceptos básicos de cada dominio.



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



El segundo objetivo es reproducir los esquemas de obtención de datos, procesamiento de la información y generación del conocimiento en las disciplinas seleccionadas mediante prácticas de laboratorio.

Contenidos mínimos:

Panorama de minería de datos y descubrimiento del conocimiento en ciencia y tecnología. Aplicaciones en astronomía, química y biología. Introducción a los sistemas de información geográficos. Aplicación de data mining en el análisis de redes de interacción y redes sociales. Componente temporal en los modelos. Visualización de datos.

○ *ENFOQUE ESTADÍSTICO DEL APRENDIZAJE*

Objetivos:

Se pretende que el alumno adquiera conocimientos que le permitan construir modelos predictivos mediante métodos de regresión múltiple y logística. Que adquiera nociones de estimación bayesiana y por máxima verosimilitud y pueda seleccionar los modelos adecuados para el tratamiento y solución correcta de los problemas tratados.

Contenidos mínimos:

Enfoques de la inferencia estadística. Estimación por Máxima Verosimilitud. Inferencia Bayesiana. Regresión lineal simple y múltiple. Estimación de mínimos cuadrados. Análisis de supuestos del modelo. Inferencia. Diagnósticos de Influencia. Multicolinealidad. Transformaciones. Variables dummies. Interacción. Métodos de selección de variables. Odds y Riesgo Relativo. Regresión logística binaria. Logits. Estimación MV. Inferencia en el modelo. Pruebas de ajuste de Hosmer y Lemeshow. Métodos de selección de variables. Medidas diagnósticas. Clasificación con Regresión logística. Regresión logística multinomial y ordinal.

Comparación de modelos y algoritmos. Evaluación y selección de modelos. LRT y devianza.



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



Indicadores de ajuste y comparación más utilizados: AIC, BIC, RMSEA, KS, medidas de parsimonia, AUC, etc. Métodos gráficos: curvas ROC, gráficos lift, gráficos de ganancia, gráficos de riesgo, etc. Factor Bayes.

Redes bayesianas. Condición de Markov. Distribuciones asociadas a una RB. Aprendizajes en una RB. Propagación de la evidencia. Aprendizaje paramétrico: MV versus Bayesiano. Redes para clasificación. Naive Bayes. Algunos tópicos a desarrollar en los trabajos de intensificación: Nociones de muestreo y remuestreo. Validación cruzada. Bootstrap, Jackknife, y métodos relacionados.

Gibbs Sampling: aplicaciones. Regresión Ridge y Lasso. Regresión por mínimos cuadrados parciales (PLSR). Tratamiento de datos faltantes.

o VISUALIZACIÓN DE LA INFORMACIÓN

Objetivos:

El objetivo consiste en impartir conocimientos conceptuales y prácticos ligados a la visualización de la información, utilizando los conocimientos ya adquiridos por los estudiantes de la Maestría en cuatrimestres anteriores. La formación y experiencia adquiridas en el curso permitirá a los alumnos estar preparados para el diseño, la implementación y la evaluación de visualizaciones interactivas o infografías.

Contenidos mínimos:

Técnicas de visualización de datos. Visualización de la Información vs. Visualización Científica. Ejemplos históricos. Modelos y marcos conceptuales: Tipos de datos (nominales, ordinales, cuantitativos, vectoriales, etc.). Datos uni- y multivaluados y multivariados. Datos no estructurados. Uso del color. Teoría del color. Aplicaciones en visualización de información. Percepción y visualización. El problema de la presentación, Foco y contexto, acercamiento semántico. Conectividad y conexiones. Visualización de grafos, redes,



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



árboles. Aplicaciones en redes sociales. Mapas Temáticos, Galaxias, Galaxias de Noticias, Mapas de Kohonen. Sistemas de Información Geográfica. Visualización Geospacial.

○ *TALLER DE TESIS II*

Objetivos:

Adquisición de conceptos de escritura científica, pasando revista a los diversos aspectos que conforman esta actividad.

Contenidos mínimos:

Objetivos y conceptos básicos. Etapas en la preparación de documentos. Metodología. Hallazgos o Resultados. Tablas. Figuras. Introducción, marco teórico y objetivos. Discusión y conclusiones. Referencias bibliográficas. Título y palabras clave. Resumen y Agradecimientos. Corrección y apreciación del manuscrito. Autoría. Criterios de elección de las revistas. Indización. Preparación y envío del manuscrito. Evaluación del manuscrito. Otros documentos académicos o científicos. Tesis de Posgrado. Ética y fraude científico. Presentaciones orales y en posters.

○ *REDES NEURONALES*

Objetivos:

El objetivo central de este curso es la presentación de las arquitecturas clásicas de Redes Neuronales aplicadas en el área de la Minería de Datos. El énfasis está puesto en la resolución de problemas de clasificación y predicción.

Tradicionalmente, las estrategias de entrenamiento han sido la herramienta más utilizada para realizar el proceso de aprendizaje de la información disponible y por tal motivo, se dedica la primera mitad del curso a este enfoque. Esto implica establecer a priori un conjunto de características referidas a la red neuronal a utilizar que condicionan la eficacia y



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



eficiencia de su funcionamiento. Como solución a este problema, la segunda mitad del curso analiza la incorporación de estrategias evolutivas aplicables al proceso de adaptación de las redes neuronales.

Contenidos Mínimos:

Inspiración biológica. Modelo de neurona. Perceptrón simple y multicapa (MLP): arquitectura, formulación matemática del algoritmo de retropropagación. Inicialización. Redes neuronales con funciones de base radial (RBF): arquitectura, fronteras de decisión, algoritmos de entrenamiento. Mapas auto-organizativos SOM y GSOM. Cuantización vectorial con aprendizaje. Algoritmos Evolutivos. Algoritmos Genéticos. Operadores. Métodos para evolución de pesos y estructura.

o *MINERÍA DE TEXTO*

Objetivos:

Presentar los fundamentos de la minería de texto, sean provenientes de la Web o de otro tipo de fuentes de información textual. Los alumnos estarán capacitados para comprender, utilizar y modificar los sistemas de "crawling", indexación, categorización y extracción de información.

Contenidos mínimos:

Introducción a Minería de Texto. Información textual. Fuentes estructuradas, semi-estructuradas y texto libre. Características de la Web. Crawling. Protocolos. Robots de navegación de la "web" ("Arañas"). Indexación y recuperación de información en la web. El modelo del Espacio Vectorial ("Vector Space Model"). Esquemas de asignación de pesos. El esquema 'tf-idf'. Asignación de un valor ('score') a los documentos. Medidas de Precisión, Recuperación y "F". Extracción de Información. Búsqueda parametrizada o por dominios. Zonas de documentos. Wrappers/Scrapers. Procesamiento de XML. Ranking basado en



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



conducta y en 'links'. Análisis de 'links'. Pagerank, pagerank por tópico. 'hubs' y autoridades, Spam. HITS. Comunidades. Clustering y Clasificación. Clasificación de textos y Métodos. Agrupamiento ("Clustering") y sus Métodos. item Minado de Opiniones, afecto o subjetividad. Web Mining: Uso. Pre-procesamiento. Análisis de logs. Análisis de hits. Análisis de visitas. Análisis de conductas.

o *DATAWAREHOUSING*

Objetivos:

Desarrollar los fundamentos básicos sobre los Sistemas de Data Warehousing. Se presentarán conceptos y técnicas de diseño utilizadas en estos sistemas para resolver aplicaciones relativas a la toma de decisiones. Mediante talleres se realizarán prácticas con herramientas de software que ofrecen soluciones a componentes de dichos sistemas.

Contenidos mínimos:

Sistemas de Data Warehouse. Conceptos Generales y Proceso de Diseño. Diseño Conceptual a partir de requerimientos o a partir de bases fuentes. Diseño Lógico Relacional. Procesos de Carga y Actualización. Técnicas de limpieza de datos. Arquitecturas de Sistemas de DW. Tecnologías de DBMS.

o *SISTEMAS DE INFORMACIÓN GEOGRÁFICA*

Objetivos:

El objetivo de este curso es introducir al alumno en las herramientas de los SIG, junto a otras que los SIG requieren como lo son los software de SIG y algunas técnicas de teledetección.

Contenidos mínimos:

Elementos de un SIG: Software, Datos, Usuarios, Información. Sistemas de coordenadas



Ministerio de Educación y Deportes
Universidad Tecnológica Nacional
Rectorado



terrestres: Geoide, Elipsoide, Proyecciones. Modelos de datos de un SIG: Raster, Vector y bases de datos relacionales. Entidades espaciales: Punto, Línea, Polígono. Raster. Georreferenciación, clasificación. Modelo Digital de Elevación s. Modelización de datos espaciales y Modelos Cartográficos.

○ *SERIES DE TIEMPO*

Objetivos:

Introducir al maestrando en los principales métodos de análisis de series temporales, tanto desde la teoría cómo con ejemplos prácticos de aplicación

Contenidos mínimos:

Definición y representación de series temporales. Representación en los dominios del tiempo y la frecuencia. Métodos de preprocesamiento: representación, indización, segmentación, visualización y medidas de similaridad. Minería de series temporales: clustering, minería de patrones, clasificación, reglas de descubrimiento, predicciones. Aplicaciones biomédicas, en negocios y economía.
