



APRUEBA CURSO DE ACTUALIZACIÓN DE POSGRADO

Buenos Aires, 3 de julio 2020

VISTO la Resolución N° 478/19 del Consejo Directivo de la Facultad Regional Paraná, a través de la cual solicita la aprobación y autorización de implementación del Curso de Actualización de Posgrado “El problema de las múltiples comparaciones y su solución en modelos predictivos” como optativo para la carrera de Maestría en Minería de Datos, y

CONSIDERANDO:

Que el curso propuesto responde a la necesidad de garantizar niveles de actualización permanente en la propuesta de formación correspondiente a la carrera de Maestría en Minería de Datos, aprobada por Ordenanza C.S. N° 1584.

Que el Consejo Superior autorizó a través de la Resolución N° 572/17 a la Facultad Regional Paraná a dictar la mencionada carrera.

Que la Comisión de Posgrado de la Universidad ha analizado los antecedentes que acompañan la solicitud y avala la presentación, y la Comisión de Ciencia, Tecnología y Posgrado recomienda 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:

ARTÍCULO 1°.- Aprobar el currículum del Curso de Actualización de Posgrado “El problema de



Ministerio de Educación
Universidad Tecnológica Nacional
Rectorado



las múltiples comparaciones y su solución en modelos predictivos” como optativo para la carrera de Maestría en Minería de Datos, que figura en el Anexo I y es parte integrante de la presente Ordenanza, en un todo de acuerdo con lo establecido por la Ordenanza N° 1313.

ARTÍCULO 2°.- Autorizar el dictado del mencionado Curso en la Facultad Regional Paraná con el cuerpo docente que figura en el Anexo II y es parte integrante de la presente Ordenanza.

ARTÍCULO 3°.- Establecer que la propuesta mencionada en el Artículo precedente quedará supeditada al cronograma de dictado de las correspondientes actividades académicas de la Facultad Regional.

ARTÍCULO 4°.- Regístrese. Comuníquese y archívese.

ORDENANZA N° 1772

UTN
DO
f.c.r.
l.p.

ING. HÉCTOR EDUARDO AIASSA
RECTOR

ING. MIGUEL ÁNGEL SOSA
Secretario General



ORDENANZA Nº 1772

ANEXO I

CURSO DE ACTUALIZACIÓN DE POSGRADO

MAESTRÍA EN MINERÍA DE DATOS

EL PROBLEMA DE LAS MÚLTIPLES COMPARACIONES Y SU SOLUCIÓN

EN MODELOS PREDICTIVOS

1. FUNDAMENTACIÓN Y JUSTIFICACIÓN

El curso se propone como optativo en la carrera Maestría en Minería de Datos. En el mismo se profundiza en la temática abordada en el curso obligatorio de la carrera denominado Aplicaciones de Minería de Datos a Economía y Finanzas, basándose en el estudio, análisis y aplicación de los modelos predictivos más avanzados, técnicas de validación de modelos y técnicas de interpretación aditivas de modelos complejos.

2. OBJETIVOS

- Conocer el problema estadístico de las múltiples comparaciones identificándolo como la verdadera y única causa del sobreajuste en la inferencia estadística (overfitting), aprender a identificarlo en los algoritmos de modelos predictivos, proveer soluciones al mismo de modo de obtener modelos no sesgados y sin sobreajuste por diseño. Aplicar dichas soluciones a las librerías existentes de árboles de decisión y gradient boosting.
- Conocer y utilizar efectivamente las técnicas “estado del arte” de algoritmos y librerías de última generación de modelado predictivo sobre datos estructurados (XGBoost,



Ministerio de Educación
Universidad Tecnológica Nacional
Rectorado



LightGBM, CatBoost) y técnicas de explicación de modelos predictivos complejos como ser los valores Shapley

- Aplicar en forma práctica a un problema concreto con datos y dimensiones reales las metodologías anteriores.

3. CONTENIDOS MÍNIMOS

Comparación de Modelos Predictivos:

El problema de las múltiples comparaciones, “look-elsewhere effect”: Distribución del máximo. Consecuencias en la inferencia estadística de la ciencia experimental. Consecuencias en los modelos predictivos

Patologías derivadas de las múltiples comparaciones en modelos predictivos: Sobreajuste (overfitting). Error de Selección. Sobre búsqueda (oversearching). Sesgo en la estimación de probabilidades

Soluciones al problema de las múltiples comparaciones: Validación Cruzada. Ajuste de Bonferroni. Randomization Testing. Atributos “Canarito”

Algoritmos de árboles no sesgados: CTree. GUIDE. QUEST. MOB

Conexiones con Curva ROC, cáscara convexa

Gradient Boosting Machines

Breve reseña de Random Forest

Algoritmo de Boosting, teoría general, algoritmo AdaBoost

Algoritmo de Gradient Boosting

Librerías: XGBoost (año 2016). LightGBM (año 2017). CatBoost (año 2018)

Regularización L1 y L2, relación con la Corrección de Laplace

Hiperparámetros y su optimización



Ministerio de Educación
Universidad Tecnológica Nacional
Rectorado



Método de Walk Forward Optimization

Cross Validation

Backtesting

Walk Forward Testing

Interpretación de modelos predictivos complejos

Métodos Globales: Dimensión de Vapnik-Chervonenkis. Gradient Boosting of Decision Stumps, su poder explicativo, su relación con la regresión logística.

Métodos locales: Shapley local additive explanations. Método LIME. Estimación de valores Shapley. Librería SHAP

4. DURACIÓN

El curso tendrá una duración de CUARENTA Y OCHO (48) horas.

5. METODOLOGÍA

El régimen de cursado previsto es presencial. El cursado prevé la combinación de clases teóricas - expositivas y actividades prácticas, particularmente, la resolución de problemas.

6. EVALUACIÓN Y PROMOCIÓN

Para la aprobación del curso se requerirá, además del 80% de asistencia, la ejecución de los trabajos prácticos y la aprobación de un examen final individual.



ORDENANZA N° 1772

ANEXO II

CURSO DE ACTUALIZACIÓN DE POSGRADO
EL PROBLEMA DE LAS MÚLTIPLES COMPARACIONES Y SU SOLUCIÓN
EN MODELOS PREDICTIVOS
MAESTRÍA EN MINERÍA DE DATOS
FACULTAD REGIONAL PARANÁ

Cuerpo Docente:

- Master Gustavo DENICOLAY PACHECO (DNI 95.614.610)
