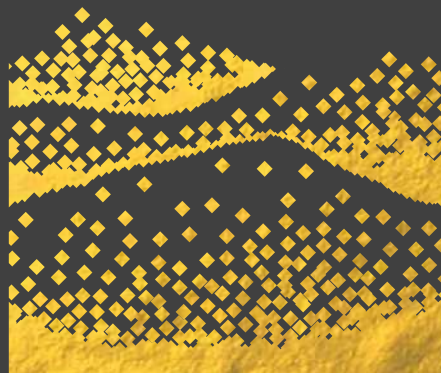


BOURBAKI

COLEGIO DE MATEMÁTICAS

Curso Intermedio

Procesamiento del Lenguaje  
Natural





## Perfil del estudiante

Los estudiantes pueden estar dedicados a cualquiera de las siguientes áreas: Ciencia de Datos, Business Intelligence, Ingeniería de Datos o Desarrollo de Software.

- I. Conocimientos previos sobre Python: Básicos.
- II. Conocimientos previos sobre Matemáticas: Álgebra Lineal y Estadística.
- III. Conocimientos previos sobre ML: Modelos de clasificación binaria y series de tiempo.



## Beneficios

El objetivo principal del curso es permitir a los Científicos de Datos en BBVA ahorrar una cantidad sustancial de tiempo en tres etapas fundamentales de las soluciones *data-driven*:

- I. El planteamiento del caso de uso tomando en cuenta la geometría, estadística y costos computacionales de la base de datos, así como la métrica adecuada del problema.
- II. Las técnicas del pre-procesamiento de los datos, la elección del algoritmo y método de calibrado que permitan reaccionar a tiempo cuando los resultados iniciales no sean los deseados por la necesidad del negocio.
- III. La defensa o crítica de un modelo matemático en términos simples y bien informados que permitan a los analistas interactuar con las distintas áreas dentro de BBVA.



# Estructura y horarios

- Número de estudiantes mínimo: 15
- BBVA tendrá acceso a las grabaciones
- Cursos en vivo con acceso a grabaciones. Once horas.
- Notas, repositorio y manuales disponibles
- Cursos entre 16:00 y 18:00 de lunes a jueves.

## Semanas de aprendizaje

- A.** Planteamiento del caso de uso y exploración de los datos.
- B.** Procesamiento, modelos matemáticos, algoritmo y métricas.
- C.** Implementación del caso de uso en Python.

## Semanas de Reto

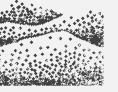
- A.** Trabajo en el reto.
- B.** Dudas & Knowledge Check.
- C.** Presentación del reto en equipos



# Retos, evaluación y acompañamiento

- 1.** Cada semana se realizará un test con 3 preguntas para revisar el aprendizaje del estudiante. Las respuestas se revisarán de manera grupal y además los estudiantes recibirán feedback personalizado.
- 2.** Al finalizar cada bloque se asignará un reto con un caso de uso semejante al implementado, los estudiantes trabajarán en equipos de tres para resolverlo y al finalizar se presentarán sus hallazgos y retroalimentación a todo el grupo.
- 3.** Tanto los profesores como los ayudantes acompañarán a los estudiantes a lo largo del curso y estarán comprometidos en resolver dudas por medio de un canal de comunicación directa. También habrán sesiones de trabajo en grupos de tres para ayudarles a los estudiantes a resolver el reto planteado.





Procesamiento del lenguaje natural:  
**sintaxis y semántica**





# Temario

## Módulo I. Análisis de sentimientos

1. Bolsa de palabras, Tf-idf & ley de *Zipf*
2. Clasificación y ranking
3. Regularización Ridge & Lasso
4. Análisis de sentimientos

## Módulo III. Redes densas y embeddings

1. Introducción a las redes densas
2. Encajes de palabras y semántica
3. Word2vec & Doc2vec
4. Aplicaciones de embeddings

## Módulo V. Grandes modelos del lenguaje

1. Soft fine-tuning con ChatGPT
2. Capas de Atención y transformer
3. BERT & GPT
4. RAG y destilación

## Módulo II. Topic Modeling

1. Modelos generativos
2. Medidas de similitud y maldición de la dimensión
3. Clustering v.s. Topic Modeling
4. Latent Dirichlet Allocation
5. *Exchangability*

## Módulo IV. Redes recurrentes y *Seq2seq*

1. Redes recurrentes y datos secuenciales
2. Redes *LSTM* y la memoria
3. Modelos *Seq2seq*
4. Resúmenes ejecutivos
5. Atención de *Bahdanau* (vainilla)



## Los profesores del curso



**Carlos Alfonso Ruiz Guido**

*Director del Colegio de Matemáticas  
Bourbaki*

Director y cofundador del Colegio de Matemáticas Bourbaki, estudió Matemáticas en la UNAM, Université d'Orsay y Oxford University. Su investigación está enfocada en la interpretabilidad y la causalidad de modelos neuronales profundos así como en la geometría de los modelos recurrentes.



**Pablo Conte**  
Profesor y Científico de Datos

Ingeniero Químico (Universidad Nacional de Mar del Plata - UNMDP, Argentina).

Especialista en Producción de Petróleo y Gas (Instituto Tecnológico de Buenos Aires - ITBA, Argentina)

Ingeniero de Reservorios y Estimulación en YPF SA (2012-2021)

Estudiante Master Science en Computación y Tecnologías Cuánticas (Democritus University of Thrace, Grecia)

Quantum Technical Member en Qualition

Quantum Software Mentee en Quantum Open Source Foundation (QOSF)



# Oferta academica



## Rudimentos de Machine Learning

(overfitting & underfitting)

Machine Learning. Regresiones lineales y árboles. Clafiscaciones Multi-clase. Clusterización y reducción de la dimensión. Percepción simple y modelos lineales. Árboles en regresión. Boosting y random Forest. K-means. t-SNE.



## Procesamiento de Lenguaje Natural

(semática y sintaxis)

Clasificación de Textos. Topic Modeling. Encajes y Redes Recurrentes. Grandes Modelos del Lenguaje. Bolsa de palabras y Tf-idf. Clustering v.s. Topic Modeling. Redes Recurrentes. BERT v.s. CHATGPT. RAG. Destilación de modelos.



## Profundización de Machine Learning

(variables latentes y secuenciales)

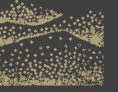
Análisis de Supervivencia. Redes Neuronales. Sistemas de Recomendación. Datos Temporales. Aprendizaje por Refuerzo. Random Survival Forest. Missing valúes. Collaborative filtering. Prophet y sus comparación con ARIMA y LSTM. Q-Learning.



## Machine Learning & AI

(Tensores y simulación)

Redes convolucionales. Redes Recurrentes. Datos Multi-Modales. Aprendizaje por refuerzo profundo. Graph Neural Networks. Modelos Seq2seq. Encajes de texto: Doc2vec. Modelos Transformer.



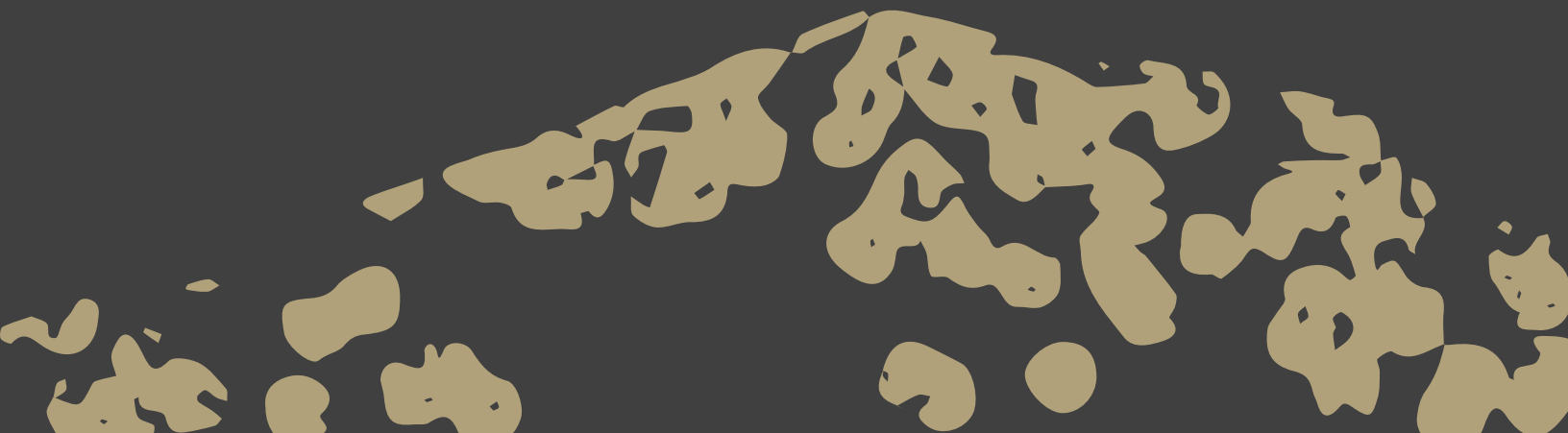
## Colegio de Matemáticas Bourbaki

---

Colegio de Matemáticas Bourbaki es un espacio para el aprendizaje personalizado. Aquí se imparten métodos de enseñanza para acercar a las personas al trasfondo matemático en múltiples fenómenos y procesos. Nuestro objetivo es vincular la academia con aplicaciones de la realidad imperante.

La institución tiene una responsabilidad con la sociedad: procurar siempre el rigor académico en todos sus servicios. Su objetivo es convertir el conocimiento matemático en una inversión sostenible, que genere riqueza.

La especialización es la piedra central de la comprensión de los problemas, el primer paso en la ruta hacia las respuestas que demanda la realidad. La selección de material educativo asertivo y la atención a los detalles correctos aleja a Colegio Bourbaki de los cursos express. Asumimos que con claridad, constancia y transparencia, se construye un camino intelectual sincero.



# BOURBAKI

ESCUELA DE MATEMÁTICAS

CURIOSIDAD Y SABER

CDMX, MEXICO

Estaremos encantados de explicar con  
detalle sobre nuestros cursos

Escríbenos por mail o WhatsApp

[info@colegio-bourbaki.com](mailto:info@colegio-bourbaki.com)

+52 56 2141 7850

escuela-  
bourbaki.com

