



التدريبية: الذكاء الاصطناعي القائم على  
- المعالجة المسبقة وهندسة الميزات الدورة  
البيانات



يونيو ٢٠٢٦ ٠٥ - ٠١



برلين

(للشخص الواحد) € ٤٩٠٠

Ref: #AI2147\_241900



## مقدمة الدورة التدريبية / لمحة عامة:



المسبقة وهندسة الميزات، التدريبية المتخصصة حول الذكاء الاصطناعي القائم يقدمه BIG BEN Training Center هذه الدورة التعلم الآلي الذين يدركون أن جودة البيانات هي وهي مصممة للمحللين، وعلماء البيانات، ومهندسي على البيانات - المعالجة المعالجة المسبقة للبيانات وموثوقة. في عصر البيانات الكبيرة (Big Data)، حجر الزاوية في بناء نماذج ذكاء اصطناعي قوية أداءً أمراً بالغ الأهمية لتحسين (Feature Engineering) وهندسة الميزات (Data Preprocessing) يُعد فهم وتطبيق تقنيات (Data Cleaning) التعامل مع القيم المفقودة (Outlier Detection) اكتشاف القيم الشاذة (Missing Values) مع القيم المفقودة في (Data Transformation) تحويل البيانات (Outlier Detection) اكتشاف القيم الشاذة (Missing Values) مع القيم المفقودة (Feature Selection) اختيار (Creation) هندسة الميزات مثل إنشاء ميزات جديدة (Feature)، بالإضافة إلى تقنيات متقدمة لضمان أن النماذج تتدرب على سيتعلم المشاركون كيفية تحضير (Dimensionality Reduction) وتقليل الأبعاد (Selection) لبناء وأداء أفضل للذكاء الاصطناعي. تهدف الدورة إلى بيانات نظيفة ومنظمة، مما يؤدي إلى تنبؤات أكثر دقة للبيانات بفعالية الاصطناعي من الألف إلى الياءة. خطوط أنابيب بيانات (Data Pipelines) قوية تدعم تزويد المشاركين بالمهارات العملية اللازمة ، الذي يؤكد دائماً على (Andrew Ng) نستلهم في هذه الدورة من أعمال البروفيسور أندرو نغ مشاريع الذكاء الذكاء



تطبيقية مكثفة ودراسات حالة من الاصطناعي"، وأن جودة هذا الوقود هي مفتاح النجاح. أن "البيانات هي الوقود الذي يشغل مجالات مختلفة، مما يعزز فهم المشاركين للجوانب تطبيقات الذكاء الاصطناعي القائم على البيانات في ستقدم الدورة أمثلة النظرية والعملية لهذه الخطوات الأساسية.



## لأ الفئات المستهدفة / هذه الدورة التدريبية مناسبة

- علماء البيانات.
- مهندسي التعلم الآلي.
- محلي البيانات.
- مهندسي البيانات.
- الباحثين في مجال الذكاء الاصطناعي.
- المطورين المهتمين بالتعلم الآلي.
- طلاب الدراسات العليا في علوم الكمبيوتر والإحصاء.
- المختصين في التحليلات التنبؤية.

## القطاعات والصناعات المستهدفة:

- تكنولوجيا المعلومات.
- الاستشارات وتحليل البيانات.
- الخدمات المالية والتأمين.
- الرعاية الصحية.
- التجارة الإلكترونية والتجزئة.
- الاتصالات.
- التصنيع.
- البحث والتطوير.
- القطاع الحكومي وما في حكمها.
- الطاقة.



## الأقسام المؤسسية المستهدفة:

- قسم علم البيانات
- قسم التعلم الآلي
- قسم التحليلات
- قسم هندسة البيانات
- قسم البحث والتطوير
- قسم الذكاء الاصطناعي
- فرق تطوير المنتجات القائمة على البيانات
- قسم إدارة البيانات

## أهداف الدورة التدريبية:

أتقن المهارات التالية: بنهاية هذه الدورة التدريبية، سيكون المتدرب قد

- الاصطناعي، فهم أهمية جودة البيانات في بناء نماذج الذكاء
- المفقودة، القدرة على تنظيف البيانات والتعامل مع القيم
- اكتشاف ومعالجة القيم الشاذة في مجموعات البيانات.
- تطبيق تقنيات تحويل وتطبيع البيانات.
- إنشاء ميزات جديدة ذات مغزى من البيانات الخام.
- اختيار الميزات الأكثر أهمية لتحسين أداء النموذج.
- الأساسية، تقليل أبعاد البيانات دون فقدان المعلومات
- بناء خطوط أنابيب للمعالجة المسبقة وهندسة الميزات.
- التعلم الآلي، تقييم تأثير المعالجة المسبقة على أداء نماذج
- بفعالية، التعامل مع البيانات النصية والفئوية والرقمية



## منهجية الدورة التدريبية:

المسبقة منهجية تدريبية عملية وموجهة نحو حل المشكلات، تهدف يعتمد BIG BEN Training Center في هذه الدورة على المنهجية محاضرات نظرية للبيانات وهندسة الميزات في سياق الذكاء الاصطناعي إلى تمكين المشاركين من إتقان المعالجة بالإضافة إلى ورش عمل تطبيقية مكثفة متعمقة حول أنواع البيانات، مشكلات الجودة، وتقنيات القائمة على البيانات. تشمل مع القيم ومكتباتها (Pandas, NumPy, Scikit-learn) Python باستخدام أدوات ولغات برمجة شائعة مثل التحويل، بيانات حقيقية. سيتم التركيز المفقودة والشاذة، وتطبيق تقنيات هندسة الميزات المشاركون بتنظيف البيانات، والتعامل البيانات والميزات على أداء نماذج التعلم الآلي. على دراسات حالة واقعية لتوضيح كيفية تأثير جودة المتقدمة على مجموعات المدربين الخبراء لضمان الخبرات وتطوير أفضل الممارسات. يتلقى المشاركون سيتم تشجيع العمل الجماعي والمناقشات لتبادل نماذج ذكاء اصطناعي قوية وموثوقة. تطوير مهاراتهم في تحضير البيانات بفعالية وبناء تغذية راجعة مفصلة من

## خريطة المحتوى التدريبي (معايير الدورة التدريبية):

### الاصطناعي. الوحدة الأولى: أساسيات البيانات ودورها في الذكاء



- أهمية البيانات النظيفة والجاهزة في التعلم الآلي.
- أنواع البيانات: رقمية، فئوية، نصية، زمنية.
- التكرار، (الضوضاء)، مشاكل جودة البيانات الشائعة (القيم المفقودة، مفهوم المعالجة المسبقة للبيانات).
- دورة حياة البيانات في مشاريع الذكاء الاصطناعي.
- (Python, Pandas) أدوات ولغات برمجة شائعة للمعالجة المسبقة.
- التفكير النقدي في جودة البيانات.

## الوحدة الثانية: تقنيات المعالجة المسبقة للبيانات.

- (Imputation) التعامل مع القيم المفقودة: الحذف، التعبئة.
- اكتشاف ومعالجة القيم الشاذة (Outliers).
- توحيد، تحجيم، تحويل البيانات (Data Transformation): تطبيع،
- (One-Hot Encoding, Label Encoding) ترميز البيانات الفئوية:
- (Stemming, Lemmatization) التعامل مع البيانات النصية: الترميز، التطهير،
- تقنيات التعامل مع البيانات الزمنية.
- تطبيق عملي على مجموعات بيانات مختلفة.

## الوحدة الثالثة: مقدمة إلى هندسة الميزات.

- مفهوم هندسة الميزات وأهميتها.
- (Creation) إنشاء ميزات جديدة من الميزات الموجودة (Feature).
- القيم المشتقة، أمثلة على الميزات المهندسة (التفاعلات، النسب،
- تقنيات التجميع (Aggregation) للميزات.
- هندسة الميزات للبيانات العددية والفئوية.
- (Embeddings) هندسة الميزات للبيانات النصية (TF-IDF, Word).
- أهمية المعرفة بالمجال في هندسة الميزات.



## الوحدة الرابعة: اختيار الميزات وتقليل الأبعاد

- وأهدافه، مفهوم اختيار الميزات ((Feature Selection
- (Methods طرق اختيار الميزات القائمة على التصفية (Filter
- (Methods طرق اختيار الميزات القائمة على التغليف (Wrapper
- (Methods طرق اختيار الميزات القائمة على التضمين (Embedded
- (Reduction مقدمة إلى تقليل الأبعاد (Dimensionality
- (PCA - Analysis تحليل المكونات الرئيسية (Principal Component
- أمثلة عملية على اختيار الميزات وتقليل الأبعاد.

## الممارسات، الوحدة الخامسة: بناء خطوط أنابيب البيانات وأفضل

- وهندسة الميزات، تصميم خطوط أنابيب (Pipelines) للمعالجة المسبقة
- أتمتة عملية تحضير البيانات
- الميزات، أفضل الممارسات في المعالجة المسبقة وهندسة
- التحديات الشائعة والأخطاء الواجب تجنبها
- (Data التعامل مع البيانات غير المتوازنة (Imbalanced
- أهمية التوثيق والتحقق من صحة البيانات
- (AutoML) مستقبل هندسة الميزات والذكاء الاصطناعي التلقائي

## الأسئلة المتكررة:

### التسجيل في الدورة؟ ما هي المؤهلات أو المتطلبات اللازمة للمشاركين قبل

لا توجد شروط مسبقة.

الإجمالي لساعات الدورة التدريبية؟ كم تستغرق مدة الجلسة اليومية، وما هو العدد



المدة إلى ٢٥٢٠- بمعدل يومي يتراوح بين ٤ إلى ٥ ساعات، تشمل فترات تمتد هذه الدورة التدريبية على مدار خمسة أيام، ساعة تدريبية، راحة وأنشطة تفاعلية، ليصل إجمالي

## سؤال للتأمل:

تضمن توفيراً مشروع ذكاء اصطناعي، كيف يمكن للمؤسسات ضمان وجود بالنظر إلى أن جودة البيانات هي مفتاح نجاح أي مستمر؟ بيانات عالية الجودة لفرق الذكاء الاصطناعي بشكل استراتيجية شاملة لإدارة البيانات

## ما الذي يميز هذه الدورة عن غيرها من الدورات؟



إغفالهما في الدورات المعالجة المسبقة للبيانات وهندسة الميزات، وهما تتميز هذه الدورة بتركيزها العميق والعملي على التقنيات مع التطبيق العملي المكثف باستخدام الأخرى. ما يميزنا هو تقديم الأسس النظرية لهذه جانبان حيويان غالباً ما يتم من مشكلات جودة يضمن أن المشاركين يكتسبون مهارات قابلة للتطبيق مجموعات بيانات حقيقية ولغات برمجة شائعة، مما تقنيات متقدمة لإنشاء واختيار الميزات التي البيانات وكيفية التعامل معها بفعالية، بالإضافة الفوري. نغطي مجموعة واسعة إلى بناء نماذج ذكاء اصطناعي تزويد المشاركين بالقدرة على تحضير البيانات تعزز أداء النماذج بشكل كبير. الدورة تركز على إلى للمهنيين الذين يرغبون في إتقان فن وعلم البيانات أكثر دقة وموثوقية. هذه الدورة هي الخيار الأمثل بفعالية، مما يؤدي لبناء حلول ذكاء اصطناعي قوية.