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

#AI2147

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

مقدمة الدورة التدريبية / لمحة عامة:

يقدم BIG BEN Training Center هذه الدورة التدريبية المتخصصة حول الذكاء الاصطناعي القائم على البيانات ـ المعالجة المسبقة وهندسة الميزات، وهي مصممة للمحللين، وعلماء البيانات، ومهندسي التعلم الآلي الذين يدركون أن جودة البيانات هي حجر الزاوية في بناء نماذج ذكاء اصطناعي قوية وموثوقة. في عصر البيانات الكبيرة (Big Data)، يُعد فهم وتطبيق تقنيات المعالجة المسبقة للبيانات (Data Preprocessing) وهندسة الميزات (Feature Engineering) أمرًا بالغ الأهمية لـتحسين أداء نماذج التعلم الآلي والذكاء الاصطناعي. ستغطي الدورة مفاهيم أساسية مثل تنظيف البيانات (Data Cleaning)، التعامل مع القيم المفقودة (Missing Values)، اكتشاف القيم الشاذة (Outlier Detection)، تحويل البيانات (Data Transformation)، بالإضافة إلى تقنيات متقدمة في هندسة الميزات مثل إنشاء ميزات جديدة (Feature Creation)، اختيار الميزات (Feature Selection)، وتقليل الأبعاد (Dimensionality Reduction). سيتعلم المشاركون كيفية تحضير البيانات بفعالية لضمان أن النماذج تتدرب على بيانات نظيفة ومنظمة، مما يؤدي إلى تنبؤات أكثر دقة وأداء أفضل للذكاء الاصطناعي. تهدف الدورة إلى تزويد المشاركين بالمهارات العملية اللازمة لـبناء خطوط أنابيب بيانات (Data Pipelines) قوية تدعم مشاريع الذكاء الاصطناعي من الألف إلى الياء. نستلهم في هذه الدورة من أعمال البروفيسور أندرو نغ (Andrew Ng)، Andrew Ng، الذي يؤكد دائمًا على أن "البيانات هي الوقود الذي يشغل الذكاء الاصطناعي"، وأن جودة هذا الوقود هي مفتاح النجاح. ستقدم الدورة أمثلة تطبيقية مكثفة ودراسات حالة من تطبيقات الذكاء الاصطناعي القائم على البيانات في مجالات مختلفة، مما يعزز فهم المشاركين للجوانب النظرية والعملية لهذه الخطوات الأساسية.

الفئات المستهدفة / هذه الدورة التدريبية مناسبة لـ:

- علماء البيانات.
- مهندسي التعلم الآلي.
- محللي 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:

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

- فهم أهمية جودة البيانات في بناء نماذج الذكاء الاصطناعي.
- القدرة على تنظيف البيانات والتعامل مع القيم المفقودة.
- 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.
- تقليل أبعاد البيانات دون فقدان المعلومات الأساسية.
- بناء خطوط أنابيب للمعالجة المسبقة وهندسة الميزات.
- تقييم تأثير المعالجة المسبقة على أداء نماذج التعلم الآلي.
- التعامل مع البيانات النصية والفئوية والرقمية بفعالية.

منهجية الدورة التدريبية:

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

خريطة المحتوى التدريبي (محاور الدورة التدريبية):

الوحدة الأولى: أساسيات البيانات ودورها في الذكاء الاصطناعي.

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

الوحدة الثانية: تقنيات المعالجة المسبقة للبيانات.

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

الوحدة الثالثة: مقدمة إلى هندسة الميزات.

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

الوحدة الرابعة: اختيار الميزات وتقليل الأبعاد.

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

الوحدة الخامسة: بناء خطوط أنابيب البيانات وأفضل الممارسات.

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

الأسئلة المتكررة:

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

لا توجد شروط مسبقة.

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

تمتد هذه الدورة التدريبية على مدار خمسة أيام، بمعدل يومي يتراوح بين 4 إلى 5 ساعات، تشمل فترات راحة وأنشطة تفاعلية، ليصل إجمالي المدة إلى 20–25 ساعة تدريبية.

سؤال للتأمل:

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

ما الذي يميز هذه الدورة عن غيرها من الدورات؟

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