

Desarrollo de Sistemas de Inteligencia Artificial

Temario de la currícula

Objetivos de Aprendizaje

- Utilizar adecuadamente las principales técnicas de búsqueda, así como la tipología de problemas donde pueden aplicarse dichas técnicas
- Resolver problemas de búsqueda atendiendo criterios de eficiencia
- Diseñar sistemas expertos y sistemas conexionistas que den respuesta a diversos problemas de clasificación, aproximación o modelización, así como a ejecutar etapas de aprendizaje sobre los mismos con el fin de prepararlos para su aplicación real

Contenidos

Bloque Fundamentos de la Inteligencia Artificial

Conceptos y definiciones. Ramas de la I.A. Aplicaciones de la Inteligencia Artificial. Técnicas

de búsqueda y resolución de problemas: Búsqueda no informada. Agente inteligente.

Estructura de un agente inteligente. Arquitectura de agentes. Búsqueda y resolución de problemas: la definición del problema. Espacio de estados. Representación en el espacio de estados. Estrategias de búsqueda. Algoritmos de búsqueda no informada. Técnicas de búsqueda y resolución de problemas: Búsqueda informada. Búsqueda primero al mejor. Búsqueda voraz.

Bloque Representación del conocimiento

Formas de representación del conocimiento: Sistemas de producción. Búsqueda e inferencia lógica: Sistemas de resolución. Encadenamiento hacia delante. Encadenamiento hacia atrás. Estrategias de Resolución

Bloque Sistemas Expertos

Definición. Arquitectura de un sistema experto. Componentes principales. Aplicaciones. Ventajas y limitaciones.

Bloque Redes Neuronales

Redes neuronales: Definición. Estructura de una red neuronal. Topologías. Campos de aplicación. Predicciones de redes neuronales. Mod

Clase 1: Introducción

¿Qué es la Inteligencia Artificial?

La Inteligencia Artificial (IA) constituye una rama de la informática dedicada al desarrollo de sistemas capaces de realizar tareas que tradicionalmente requieren inteligencia humana. Estos sistemas procesan información, aprenden de la experiencia, se adaptan a nuevas situaciones y resuelven problemas complejos. A diferencia de la programación convencional, donde cada instrucción debe ser explícitamente codificada, la IA permite que las máquinas adquieran conocimientos y habilidades a través del aprendizaje. Su objetivo fundamental es desarrollar algoritmos que permitan a las computadoras razonar, percibir, comprender y actuar de manera autónoma en diversos entornos y contextos, simulando capacidades cognitivas humanas como el aprendizaje, la planificación y la creatividad.

Fases de la Inteligencia Artificial

Inteligencia Artificial

La Inteligencia Artificial Simbólica representa la primera generación de sistemas inteligentes, basada en la manipulación de símbolos y reglas lógicas. Estos sistemas utilizan representaciones explícitas del conocimiento y mecanismos de inferencia para resolver problemas. Ejemplos característicos son los sistemas expertos que emulan el razonamiento de especialistas humanos en dominios específicos mediante bases de conocimiento y motores de inferencia. Esta aproximación se fundamenta en la hipótesis de que la inteligencia humana puede modelarse como procesamiento simbólico, permitiendo razonamiento lógico y manejo de conceptos abstractos, aunque muestra limitaciones al enfrentarse a problemas de percepción y aprendizaje en entornos cambiantes.

Machine Learning

El Machine Learning constituye la segunda fase evolutiva de la IA, donde los sistemas adquieren la capacidad de aprender patrones a partir de datos sin ser explícitamente programados para cada tarea. Esta aproximación se basa en algoritmos estadísticos que permiten a las máquinas identificar

regularidades y relaciones en los datos. El aprendizaje puede ser supervisado (con ejemplos etiquetados), no supervisado (descubriendo estructuras inherentes) o por refuerzo (mediante prueba y error con recompensas). Técnicas como regresión, árboles de decisión, agrupamiento y clasificación permiten a los sistemas generalizar a partir de ejemplos, adaptándose a nuevos datos y mejorando su rendimiento con la experiencia.

Redes Neuronales Artificiales

Las Redes Neuronales Artificiales representan un paradigma inspirado en la estructura y funcionamiento del cerebro humano. Consisten en unidades de procesamiento interconectadas (neuronas artificiales) organizadas en capas, que transforman señales de entrada en salidas mediante pesos sinápticos ajustables. Cada neurona recibe entradas ponderadas, aplica una función de activación y genera una respuesta. Durante el entrenamiento, estos sistemas modifican sus conexiones mediante algoritmos como la retropropagación, minimizando la diferencia entre las salidas producidas y las deseadas. Su capacidad para modelar relaciones complejas y no lineales las hace idóneas para reconocimiento de patrones, clasificación y aproximación de funciones en diversos dominios.

Deep Learning

El Deep Learning constituye la evolución más reciente del Machine Learning, caracterizada por el uso de redes neuronales con múltiples capas ocultas que permiten el aprendizaje de representaciones jerárquicas. Estas arquitecturas profundas capturan progresivamente características de mayor abstracción y complejidad, desde elementos básicos hasta conceptos sofisticados. Modelos como redes convolucionales (CNN), redes recurrentes (RNN) y transformers han revolucionado campos como visión por computadora, procesamiento de lenguaje natural y generación de contenido. Su extraordinaria capacidad para procesar datos no estructurados ha permitido avances significativos en reconocimiento de imágenes, traducción automática, síntesis de voz y otras tareas previamente inabordables, aunque requieren grandes volúmenes de datos y recursos computacionales considerables.

Aplicaciones y Perspectivas

La Inteligencia Artificial ha trascendido el ámbito académico para transformar numerosos sectores como la medicina, finanzas, transporte y entretenimiento. Los sistemas de IA asisten en diagnósticos médicos, optimizan operaciones empresariales, potencian vehículos autónomos y generan contenido creativo. Sin embargo, su implementación conlleva desafíos técnicos y éticos significativos, desde la interpretabilidad de los modelos hasta consideraciones de privacidad, sesgos y equidad. El futuro de la IA apunta hacia sistemas más robustos, explicables y alineados con valores humanos, integrando diferentes paradigmas para desarrollar inteligencia más general y versátil. La convergencia con otras tecnologías como computación cuántica, internet de las cosas y robótica promete expandir aún más el horizonte de posibilidades de esta disciplina en constante evolución.

Clase 2: Industria 4.0 y el Origen de la IA

Las Revoluciones Industriales

Primera Revolución Industrial (1760-1840): Marcó el paso de la producción manual a la mecanizada. Surgió en Inglaterra con la introducción de la máquina de vapor y la mecanización de la industria textil. El carbón se convirtió en la principal fuente de energía, transformando los sistemas de transporte, manufactura y comunicación.

Segunda Revolución Industrial (1870-1914): Caracterizada por la producción en masa y la línea de montaje. El acero, la electricidad y el petróleo impulsaron nuevas industrias. Surgieron el teléfono, el automóvil y los avances químicos. Esta etapa vio el desarrollo de grandes corporaciones y la organización científica del trabajo.

Tercera Revolución Industrial (1960-2000): También llamada revolución digital, se distinguió por la automatización, los semiconductores y la informática. Internet transformó las comunicaciones globales mientras los ordenadores personales revolucionaron el trabajo y el acceso a la información, sentando las bases para la globalización moderna.

Cuarta Revolución Industrial (2000-actualidad): Fusiona tecnologías físicas, digitales y biológicas. Se caracteriza por la inteligencia artificial, robótica avanzada, internet de las cosas, computación cuántica, biotecnología y nanotecnología. Está difuminando las fronteras entre lo físico, digital y biológico, transformando radicalmente industrias, economías y sociedades.

El término "Industria 4.0" se originó en Alemania como parte de una iniciativa gubernamental en 2011 para modernizar y mejorar la competitividad de la industria manufacturera del país mediante la implementación de tecnologías avanzadas.

La Cuarta Revolución Industrial

La Cuarta Revolución Industrial representa una transformación fundamental en la manera en que vivimos, trabajamos y nos relacionamos, impulsada por la convergencia de sistemas digitales, físicos y biológicos. A diferencia de las revoluciones industriales anteriores, esta se caracteriza por su velocidad exponencial, alcance global y el impacto sistémico en todas las industrias y sociedades. Esta nueva era tecnológica está reconfigurando gobiernos, empresas, educación y la vida cotidiana a una escala sin precedentes.

Internet de las Cosas (IoT)

El Internet de las Cosas constituye una red de objetos físicos interconectados que recopilan e intercambian datos. Estos dispositivos, equipados con sensores, software y otras tecnologías, pueden comunicarse entre sí y con sistemas más amplios sin intervención humana. El IoT ha transformado industrias como la manufactura (con fábricas inteligentes), la salud (mediante dispositivos médicos conectados), la agricultura (a través de sensores de cultivos y sistemas de riego automatizados) y las ciudades (con infraestructuras inteligentes para gestión de tráfico, energía y residuos).

La verdadera potencia del IoT radica en su capacidad para generar cantidades masivas de datos que, cuando se procesan y analizan, proporcionan información valiosa para mejorar procesos,

anticipar necesidades y optimizar recursos. Esta conectividad ubicua está creando ecosistemas más eficientes y adaptables.

Inteligencia Artificial (IA)

La Inteligencia Artificial comprende sistemas computacionales capaces de realizar tareas que normalmente requieren inteligencia humana. Estos sistemas pueden percibir su entorno, razonar, aprender, adaptarse y actuar de manera autónoma o semi-autónoma. La IA abarca diversas técnicas como el aprendizaje automático, el aprendizaje profundo, el procesamiento del lenguaje natural y la visión por computadora.

En el contexto de la Cuarta Revolución Industrial, la IA está impulsando la automatización inteligente de procesos industriales, mejorando la toma de decisiones empresariales mediante análisis predictivos, personalizando productos y servicios, y habilitando sistemas autónomos como vehículos sin conductor y robots colaborativos. Su integración con otras tecnologías como IoT multiplica su impacto, permitiendo sistemas más inteligentes y adaptativos.

Internet de los Datos (IoD)

El Internet de los Datos se refiere a la inmensa infraestructura y ecosistema destinados a la recolección, almacenamiento, procesamiento, análisis y distribución de datos a escala masiva. Va más allá del simple almacenamiento para incluir todo el ciclo de vida de los datos: desde su generación (a menudo por dispositivos IoT) hasta su transformación en conocimiento accionable.

El IoD incorpora tecnologías como big data, computación en la nube, edge computing y lagos de datos para gestionar volúmenes sin precedentes de información estructurada y no estructurada. Esta capa de datos conecta las demás tecnologías de la Cuarta Revolución Industrial, proporcionando la materia prima necesaria para algoritmos de IA, aplicaciones industriales y servicios inteligentes. El valor estratégico del IoD radica en su capacidad para convertir datos en insights que impulsan la innovación y nuevos modelos de negocio.

Internet de los Servicios (IoS)

El Internet de los Servicios representa la evolución de la computación en la nube hacia un paradigma donde cualquier capacidad puede ofrecerse como servicio a través de internet. Engloba desde infraestructura básica (IaaS), plataformas de desarrollo (PaaS), software (SaaS), hasta funciones más específicas como inteligencia artificial, blockchain o análisis de datos ofrecidos como servicios modulares y escalables.

El IoS permite a organizaciones y usuarios acceder a tecnologías avanzadas sin grandes inversiones iniciales, democratizando el acceso a capacidades que antes estaban reservadas a grandes corporaciones. Esta arquitectura de servicios interconectados facilita la creación de aplicaciones compuestas y ecosistemas digitales complejos, acelerando la innovación y permitiendo modelos de negocio más ágiles y adaptables a las condiciones cambiantes del mercado.

Internet de las Personas (IoP)

El Internet de las Personas va más allá de la conectividad entre dispositivos para centrarse en las conexiones humanas mediadas por tecnología. Abarca las redes sociales, plataformas colaborativas, tecnologías wearable, interfaces cerebro-máquina y sistemas de realidad aumentada/virtual que facilitan interacciones más naturales y personalizadas con el mundo digital.

El IoP reconoce al ser humano como el centro del ecosistema tecnológico, buscando mejorar la experiencia, productividad y bienestar personal a través de interfaces intuitivas y tecnologías adaptativas. Esta dimensión humana de la Cuarta Revolución Industrial está transformando el trabajo (con colaboración remota y equipos híbridos), la educación (mediante aprendizaje personalizado), la salud (a través de medicina de precisión) y el ocio (con experiencias inmersivas).

Convergencia y Sinergia Tecnológica

La verdadera revolución no surge de estas tecnologías de forma aislada, sino de su integración sinérgica. Cuando los dispositivos IoT generan datos (IoD) que alimentan algoritmos de IA, que a su vez se implementan a través de servicios en la nube (IoS) para mejorar experiencias humanas (IoP), emerge un ecosistema tecnológico unificado con capacidades transformadoras.

Esta convergencia está generando nuevos modelos de negocio, reconfigurando cadenas de valor y difuminando las fronteras entre industrias tradicionales. Las empresas más innovadoras no se limitan a adoptar tecnologías individuales, sino que desarrollan estrategias holísticas que aprovechan estas interconexiones para crear valor de formas previamente imposibles.

Clase 4: Tipos de IA/Introducción a Python

Bloque 1: Tipos de IA

Clasificación de la IA

Por Capacidad: Débil (ANI), General (AGI) y Super Inteligencia (ASI).

Por Función: Aprendizaje Automático, Aprendizaje Profundo, Procesamiento de Lenguaje Natural, Visión por Computadora, Sistemas Expertos, Sistemas de Control y Robótica.

Por Enfoque: Simbólica, Conexionista, Evolutiva, Basada en Reglas, Basada en Agentes y Híbrida.

Por Sector: Industria, Salud, Transporte, Finanzas, Comercio, etc.

Por Nivel de Autonomía: Autónoma, Semi-autónoma y Controlada.

Por Tipo de Dato: Datos estructurados, Datos no estructurados y Datos semiestructurados.

NLP

El Natural Language Processing es una rama de la inteligencia artificial que se enfoca en la interacción entre los ordenadores y los seres humanos a través del lenguaje natural.

Implica una serie de técnicas de aprendizaje automático y lingüística computacional, como el análisis semántico, la desambiguación del sentido de las palabras, la clasificación de textos y la generación de lenguaje natural.

Tiene como objetivo permitir que los ordenadores comprendan, interpreten y generen lenguaje humano de manera natural, de tal forma que se puedan realizar tareas como el reconocimiento de voz, la traducción automática, el análisis de sentimientos, la generación de texto, el etiquetado de texto, etc.

Se apoya en redes CNN y RNN para resolver problemas relacionados con el lenguaje humano, como el reconocimiento de voz, comprensión del lenguaje natural, generación del lenguaje natural, traducción automática, clasificación del texto, extracción de información.

Las estrategias que usa son tokenización, part of speech, shallow parsing o chunks, word2vec.

Algunos de los usos más comunes de este tipo de algoritmo incluyen Resumen de textos
ChatBots, generación de Keywords. generación de Textos, reconocimiento de Entidades
Traducción de Idiomas, clasificación Automática de Textos

Chatbot

Son sistemas de inteligencia artificial diseñados para simular una conversación humana.

Pueden ser programados para realizar varios tipos de tareas, desde responder preguntas sencillas hasta hacer reservaciones o compras en línea.

Además, pueden integrarse en diferentes plataformas, para interactuar con los usuarios de manera eficiente.

Sistemas Expertos

Son programas de ordenador que imitan el razonamiento y la toma de decisiones de un experto humano en un campo específico, como la medicina, la ingeniería, la finanzas o la informática, etc.

Estos sistemas utilizan una base de conocimientos y reglas específicas para un determinado dominio, y un motor de inferencia que permite al sistema hacer preguntas, analizar datos y llegar a conclusiones.

Los sistemas expertos también suelen tener una interfaz de usuario que permite a los usuarios interactuar con el sistema y obtener respuestas y recomendaciones.

Se pueden dividir en varias ramas, dependiendo de su enfoque y funcionalidad específicos. Algunas de las más comunes incluyen: sistemas basados en Reglas, sistemas basados en Casos, sistemas basados en Redes Neuronales, sistemas basados en Lógica Difusa y sistemas basados en Ontologías

Sistemas de Control

Son sistemas diseñados para controlar y regular el comportamiento de otros sistemas o procesos, con el fin de mantener un cierto nivel de desempeño o alcanzar un objetivo específico.

En los sistemas de control, se mide el estado actual del sistema o proceso, se compara con el estado deseado y se ajusta el comportamiento del sistema o proceso mediante el feedback para alcanzar el objetivo deseado.

Pueden ser manuales, semiautomáticos o completamente automatizados, y pueden utilizar diferentes técnicas de control.

Se pueden dividir en varias ramas, dependiendo de su enfoque y funcionalidad específicos:

Control Clásico, Control Digital, Control Adaptativo, Control Óptimo, Control de Retroalimentación

Robótica

Incluye diferentes áreas de conocimiento, como la mecánica, la electrónica, la informática y la inteligencia artificial.

Desde el punto de vista del desarrollo de la inteligencia artificial, la robótica se enfoca en la creación de robots inteligentes y autónomos que puedan aprender y adaptarse a diferentes situaciones y entornos.

La IA se utiliza en la robótica para permitir que los robots tengan la capacidad de tomar decisiones y realizar tareas complejas de manera autónoma, utilizando diferentes tipos de algoritmos

Dentro de la robótica, la IA puede enfocarse en resolver una multitud de problemáticas diferentes: navegación y Localización, manipulación y Agarre, reconocimiento y Seguimiento de Objetos, Aprendizaje y Adaptación, colaboración Humano-Robot.

Clase 5: Analisis EDA

Exploratory Data Analysis

El Análisis Exploratorio de Datos o Exploratory Data Analysis, tiene como finalidad examinar los datos previamente a la aplicación de cualquier técnica estadística.

De esta forma se consigue un entendimiento básico de los datos y de las relaciones existentes entre las variables analizadas.

¿Qué hace el EDA?

El EDA proporciona métodos sencillos para organizar y preparar los datos, detectar fallos en el diseño y recogida de datos, tratamiento y evaluación de datos ausentes, identificación de casos atípicos, etc.

¿Qué nos ayuda a entender?

Algunas de las preguntas que podemos responder gracias a realizar un EDA, son las siguientes:

¿Existe algún sesgo en los datos recogidos? ¿Hay errores en la codificación de los datos?

¿Cómo se sintetiza y presenta la información contenida en un conjunto de datos?

¿Existen datos atípicos? ¿Cuáles son? ¿Cómo tratarlos?

¿Hay datos ausentes? ¿Tienen algún patrón sistemático? ¿Cómo tratarlos?

Etapas del EDA

Un análisis EDA comprende varias etapas, las cuales no siempre se aplican de forma lineal tales como preparar los Datos para hacerlos accesibles, realizar un análisis gráfico y estadístico de la naturaleza de las variables, analizar correlaciones, evaluar supuestos distribucionales y asimetría, identificar posibles Outliers, identificar Missing Values

Clase 6: Áreas de aplicación de la IA

Finanzas

Análisis de riesgo crediticio: Los algoritmos de machine learning pueden analizar grandes cantidades de datos para evaluar el riesgo crediticio de los solicitantes de préstamos.

Detección de fraudes: cada vez es más común el uso de IA para detectar patrones de actividad sospechosa en las transacciones financieras y alertar a los equipos de seguridad.

Pronóstico de precios: se pueden aplicar algoritmos para predecir los precios de los activos financieros, como las acciones, los bonos y los productos básicos.

Retail

Recomendación de productos personalizados: Los sistemas de inteligencia artificial se utilizan para analizar los datos de compra de los clientes y recomendar productos personalizados.

Gestión de inventario: La inteligencia artificial se utiliza para optimizar la gestión de inventario y reducir los costos de almacenamiento.

Análisis de precios: Los algoritmos de inteligencia artificial se utilizan para analizar los precios de los productos y ajustarlos en función de factores como la demanda y la competencia.

Publicidad y Marketing

Publicidad programática: La inteligencia artificial se utiliza para optimizar la colocación de anuncios y maximizar el retorno de la inversión publicitaria.

Análisis de sentimiento de las redes sociales: La inteligencia artificial se utiliza para analizar los comentarios y publicaciones de las redes sociales y detectar patrones de sentimiento hacia una marca o un producto.

Personalización de campañas de marketing: Los sistemas de inteligencia artificial se utilizan para analizar los datos de los clientes y crear campañas de marketing personalizadas y relevantes.

Recursos Humanos: Selección automatizada de candidatos: La inteligencia artificial se utiliza para analizar los currículums y las habilidades de los candidatos y seleccionar a los más adecuados para un puesto de trabajo.

Análisis de la satisfacción de los empleados: Los sistemas de inteligencia artificial se utilizan para analizar las encuestas y otros datos para comprender la satisfacción y el compromiso de los empleados.

Identificación de patrones de rotación de personal: Los algoritmos de inteligencia artificial se utilizan para identificar patrones en los datos de rotación de personal y determinar las causas subyacentes de la rotación.

Educación

Tutoriales personalizados: Los sistemas de IA pueden adaptarse a las necesidades y habilidades de los estudiantes para ofrecer tutoriales personalizados.

Asistencia en la tarea: Los chatbots pueden ser utilizados para ayudar a los estudiantes a resolver preguntas o problemas.

Aprendizaje adaptativo: Los sistemas de inteligencia artificial pueden analizar el desempeño de los estudiantes en tiempo real y ajustar el nivel de dificultad de las tareas en consecuencia.

Salud

Diagnóstico médico: Los sistemas de inteligencia artificial pueden analizar grandes cantidades de datos médicos y ofrecer diagnósticos más precisos y rápidos.

Telemedicina: se puede utilizar para permitir consultas médicas virtuales, lo que puede mejorar la accesibilidad de los servicios de salud.

Monitorización del estado de salud: Los dispositivos de seguimiento de salud, como los relojes inteligentes, pueden utilizar algoritmos de inteligencia artificial para detectar patrones en los datos de seguimiento y alertar al usuario si se detecta algún problema.

Transporte

Sistemas de navegación: Los sistemas de navegación de los vehículos pueden utilizar algoritmos de inteligencia artificial para ofrecer rutas más eficientes y predecir el tráfico.

Vehículos autónomos: La inteligencia artificial es fundamental para la conducción autónoma de vehículos.

Gestión del tráfico: Los sistemas de inteligencia artificial pueden analizar los patrones de tráfico y ajustar los semáforos para mejorar la fluidez del tráfico.

Agricultura

Monitorización del clima: Los sistemas de inteligencia artificial se utilizan para analizar los datos meteorológicos y predecir el clima, lo que puede ayudar a los agricultores a planificar la siembra y la cosecha.

Detección de enfermedades de las plantas: La inteligencia artificial se utiliza para detectar enfermedades en las plantas mediante el análisis de imágenes de las hojas.

Predicción de rendimientos: Los algoritmos de IA se utilizan para predecir los rendimientos de los cultivos y ajustar el uso de los recursos en consecuencia.

Energía

Predicción de la demanda energética: Los modelos de machine learning se utilizan para predecir la demanda de energía en función de factores como la hora del día, la temperatura y los días festivos.

Optimización de la generación de energía renovable: La inteligencia artificial se utiliza para optimizar la producción de energía renovable, como la energía solar y eólica, teniendo en cuenta factores como la intensidad del viento y la luz solar.

Análisis de datos de sensores: Los sistemas de inteligencia artificial se utilizan para analizar los datos de los sensores en las centrales eléctricas y detectar problemas antes de que ocurran fallos.

Manufactura

Mantenimiento predictivo: Los algoritmos de inteligencia artificial se utilizan para predecir el fallo de la maquinaria y programar el mantenimiento antes de que ocurra el fallo.

Optimización de la cadena de suministro: La inteligencia artificial se utiliza para optimizar la cadena de suministro, teniendo en cuenta factores como la demanda, la disponibilidad de materiales y la capacidad de producción.

Inspección automatizada de calidad: La inteligencia artificial se utiliza para inspeccionar los productos fabricados y detectar defectos antes de que salgan de la línea de producción.

Medios de Comunicación

Recomendación de contenidos: Los sistemas de inteligencia artificial se utilizan para ofrecer recomendaciones personalizadas de contenidos en función de los intereses de los usuarios.

Análisis de sentimiento de las redes sociales: La inteligencia artificial se utiliza para analizar los comentarios y publicaciones de las redes sociales y detectar patrones de sentimiento hacia una marca o un producto.

Creación automatizada de contenidos: La inteligencia artificial se utiliza para crear contenido generado por ordenador, como resúmenes deportivos y noticias, a partir del análisis de grandes cantidades de datos.

Clase 7: Análisis Estadístico Univariado, Bivariado y Multivariado

Análisis estadístico

Es la ciencia de recopilar, explorar y presentar grandes cantidades de datos para descubrir patrones y tendencias subyacentes.

Las estadísticas se aplican todos los días, en investigación, industria y gobierno, para volverse más científicas sobre las decisiones que deben tomarse.

Tipos de Análisis estadístico

En el mundo de la Estadística, se distinguen tres tipos de análisis principales según el número de variables que se analicen conjuntamente tales como Univariado, Bivariado y Multivariado

Análisis univariado

Consiste en el análisis de cada una de las variables estudiadas por separado, se basa exclusivamente en una única variable.

Es el análisis más básico y por ende más primario. Se considera un análisis de tipo Descriptivo y no relacional o de causalidad

Las técnicas más frecuentes de análisis univariado son la distribución de frecuencias para una tabla univariada y el análisis de las medidas de tendencia central de la variable (media, mediana, moda, varianza , desviación estándar, cuartiles entre otros).

Análisis Bivariado

Es una de las formas más simples de análisis estadístico, que se utiliza para averiguar si existe una relación entre dos conjuntos de valores. Por lo general involucra las variables X e Y.

El análisis bivariado no es lo mismo que el análisis de datos de dos muestras. Con el análisis de datos de dos muestras, X e Y no están directamente relacionadas.

Ahora bien, cuando empleamos dos o más variables en nuestros estudios o investigaciones, buscando conocer causalidad, efectos o correlaciones podemos hablar de análisis bivariados o multivariados.

Un análisis bivariado típico puede ser por ejemplo, determinar la relación entre la ingesta de carbohidratos en el peso de los adolescentes. Decimos que se trata de un análisis bivariado justamente porque estamos analizamos dos dimensiones:

Una variable dependiente que viene a ser el peso de la población analizada.

La variable independiente que será la cantidad o ingesta de carbohidratos en su dieta.

Análisis Multivariado

Es una rama de la estadística que abarca la observación y el análisis simultáneos de más de una variable respuesta.

Los estudios multivariados son similares a los univariados y bivariados con la diferencia de que tienen más de dos variables independientes y en algunos casos múltiples variables dependientes.

Se emplean estudios multivariantes cuando requieren examinar la relación entre múltiples factores al mismo tiempo. Se diferencia claramente de los estudios univariados y bivariados en que plantean más de una variable dependiente y varias independientes.

Objetivos del Análisis Multivariado

El objetivo del análisis multivariado es variable en relación a lo que queremos conseguir como optimizar los datos o simplificar la estructura, ordenar y agrupar, investigar la relación de dependencia entre variables, relación predictiva entre variables, construcción y prueba de hipótesis.

Clase 9: Entorno del Desarrollo de Sistemas de IA

Roles dentro del mundo de los Datos

Análisis de datos: Data Analyst, Business Analyst, Data Engineer

Ciencia de datos: Data Scientist, Machine Learning Engineer, Deep Learning Engineer, Data Analyst (Avanzado)

Visualización de datos: Data Visualization Specialist, Business Intelligence Analyst, Dashboard Designer, Infographics Designer

Gestión de datos: Data Manager, Data Architect, Database Administrator, ETL Developer, Data Governance Analyst, Data Engineer

Minería de datos: Data Miner, Text Mining Specialist, Web Mining Specialist, Image Processing Engineer

Big Data: Big Data Engineer, Hadoop Developer, Spark Developer, NoSQL Developer, Data Pipeline Engineer

Gestión de Datos

¿Qué es un Data Warehouse?

Es un repositorio centralizado de datos que se utiliza para almacenar, consolidar y gestionar datos de diversas fuentes con el fin de facilitar el análisis de negocios y la generación de informes.

Ventajas: Proporciona una visión única y consistente de los datos. Mejora el rendimiento de las consultas para análisis complejos. Facilita la generación de informes y paneles de control.

Desventajas: Requiere una estructura de datos predefinida. No es adecuado para el almacenamiento de datos no estructurados o semiestructurados. Puede ser costoso de implementar y mantener.

¿Qué es un Data Lake?

Es un repositorio de almacenamiento escalable que puede contener datos en bruto y sin procesar de diversas fuentes en su formato original. No impone una estructura de datos rígida

Ventajas:

Acomoda una amplia variedad de datos, incluidos datos no estructurados y semiestructurados. Escalabilidad para grandes volúmenes de datos. Flexibilidad para explorar y analizar datos de manera más libre.

Desventajas: Puede ser difícil de gestionar y mantener. El proceso de transformación y limpieza de datos puede ser más complejo en comparación con un Data Warehouse.

Menos adecuado para consultas de alto rendimiento.

¿Qué es un Data Fabric?

Es un enfoque que busca conectar y gestionar datos de manera cohesiva en una organización, independientemente de dónde se almacenen los datos. Proporciona una capa unificada de acceso y gestión de datos.

Ventajas: Permite la integración y acceso uniforme a datos distribuidos. Facilita la colaboración entre equipos y aplicaciones. Mejora la visibilidad y el gobierno de los datos.

Desventajas: Requiere una planificación y arquitectura adecuadas para su implementación. Puede requerir inversiones significativas en tecnología y formación.

¿Qué es un Data Mesh?

Es un enfoque emergente que busca descentralizar la gestión de datos al tratar los datos como productos y distribuir la responsabilidad de los datos en toda la organización.

Ventajas: Escalabilidad al distribuir la gestión de datos entre equipos. Promueve la propiedad y responsabilidad de los datos. Fomenta la agilidad y la innovación en el uso de datos.

Desventajas: Puede requerir un cambio cultural significativo en la organización. La implementación puede ser compleja y requerir una planificación cuidadosa. La gobernanza de datos puede ser un desafío si no se gestiona adecuadamente.

¿Qué es un Data Lake?

Es una solución de almacenamiento de datos que permite a las empresas almacenar grandes cantidades de datos estructurados, semiestructurados y no estructurados en un solo lugar.

A diferencia de las soluciones de almacenamiento de datos tradicionales, como los almacenes de datos, que requieren la definición de esquemas de datos y la transformación de datos antes de ser almacenados, un Data Lake almacena datos en su formato original y permite su procesamiento en una etapa posterior.

Ventajas de un Data Lake

Un Data Lake puede mejorar el acceso y la calidad de los datos para el desarrollo de IA de varias maneras: Almacena datos de diferentes fuentes y formatos, lo que permite un acceso amplio para entrenar modelos de IA. Almacenar datos en su formato original reduce la necesidad de limpieza y transformación, ahorrando tiempo y recursos en el procesamiento de datos para modelos de IA. Mejora la calidad de los datos y permite descubrir patrones y tendencias a través de análisis exploratorios más detallados.

Claves para Implementar un Data Lake

Definir un plan claro para el almacenamiento de datos y el acceso a los mismos.

Evaluar y seleccionar tecnologías adecuadas para el almacenamiento y procesamiento de datos. Diseñar un esquema de metadatos claro y preciso que describa los datos almacenados. Establecer políticas de acceso y seguridad para proteger los datos almacenados. Establecer prácticas de mantenimiento de datos que aseguren la calidad y la integridad de los mismos. Automatizar procesos y tareas rutinarias para mejorar la eficiencia en la gestión del Data Lake.

Clase 10: Revisión de Pares

La revisión de pares es una práctica común en la industria de la programación que consiste en que dos o más programadores revisen el código de uno de ellos para encontrar errores y mejorar la calidad del mismo.

Esta práctica es aplicable, no solo a programación y código, sino que también puede ser aplicada para tener diferentes perspectivas de un análisis, o del desarrollo de un proyecto.

Beneficios de la revisión de Pares: aumenta la calidad del código, mejora la detección temprana de errores, incrementa el conocimiento compartido entre los miembros del equipo, fomenta la colaboración y el trabajo en equipo.

Métodos para la Revisión de Pares

Revisión Cara a Cara: se lleva a cabo en reuniones presenciales donde los programadores discuten el código en persona.

Revisión en Línea: se realiza mediante herramientas en línea que permiten compartir el código para que sea revisado y comentado por otros programadores.*

Revisión Formal: se trata de una revisión estructurada que sigue un proceso formal y predefinido.

¿Qué debemos tomar en cuenta?

- Establecer reglas y normas claras para la revisión
- Ser constructivo y respetuoso en los comentarios
- Centrarse en los errores importantes y no en los detalles menores
- Evitar el sesgo y mantener la objetividad
- Ser puntual en la revisión y la corrección

Clase 11: Modelos de IA Generativos

IA Generativa

Es una rama de la inteligencia artificial que se enfoca en crear algoritmos que tienen la capacidad de generar información, de manera tal que parezca creada por seres humanos.

La idea es que estos algoritmos puedan producir contenido original de manera autónoma, sin la necesidad de que un programador humano les diga exactamente qué hacer.

Tipos de IA Generativa

Redes Generativas Adversariales (GANs): son un tipo de red neuronal que consiste en dos partes: un generador y un discriminador. El generador crea nuevas muestras de datos, mientras que el discriminador evalúa si las muestras son auténticas o generadas.

Modelos Generativos Autoregresivos: estos modelos aprenden a generar secuencias de datos (como texto o música) al predecir la siguiente entrada en una secuencia. El modelo comienza con una entrada inicial (por ejemplo, una palabra o una nota musical) y luego genera secuencias basadas en las predicciones que hace.

Modelos Generativos de Flujo: estos modelos transforman datos de una distribución conocida a otra distribución desconocida, a menudo creando muestras en un espacio de alta dimensionalidad.

Ejemplos de IA Generativas

GPT: para generar texto.

DALL-E 2: para generar imágenes y arte a partir de una descripción de texto.

Midjourney: para generar imágenes realistas.

Stable Diffusion: para generar imágenes realistas.

BARD: para generar texto.

GitHub Copilot: para sugerir código.

CLIP: para clasificar y buscar imágenes.

Amper Music: para generar música.

Synthesia: para generar videos.

Pros de las IA Generativas

Creatividad: Tienen la capacidad de generar contenido nuevo y original.

Productividad: Permite automatizar o simplificar tareas.

Personalización: pueden ser entrenadas para producir contenido personalizado y específico para cada usuario.

Predicción: pueden ser utilizadas para predecir patrones y tendencias en grandes conjuntos de datos.

Avance en la ciencia: pueden ser utilizadas en campos para acelerar el descubrimiento de nuevos compuestos o procesos.

Contras de las IA Generativas

Sesgo: pueden perpetuar sesgos sociales y culturales existentes.

Falta de control: En algunos casos, pueden generar contenido que no se ajusta a los estándares de calidad o éticos.

Aprendizaje insuficiente: requieren grandes cantidades de datos de entrenamiento para funcionar correctamente.

Costo: La construcción y el mantenimiento de modelos generativos de alta calidad puede ser costoso.

Desplazamiento laboral: La automatización de tareas creativas puede llevar a la eliminación de trabajos que antes eran realizados por humanos.

Desafíos de las IA Generativas

Uno de los desafíos más importantes a los que se enfrentan estas IA es el de la interpretación y evaluación de la calidad del contenido generado.

Esto se debe a que, a menudo, la generación de contenido es un proceso subjetivo que depende de los gustos y preferencias de las personas.

Por lo tanto, es necesario contar con herramientas y métricas que permitan evaluar la calidad del contenido generado de manera objetiva.

Otro desafío importante es el de la interpretación de los resultados generados por las IA generativas.

En muchos casos, los modelos generativos pueden ser difíciles de interpretar, lo que dificulta su uso en aplicaciones que requieren explicabilidad y transparencia.

En consecuencia, es necesario continuar investigando técnicas y herramientas que permitan interpretar y comprender los resultados generados por las IA generativas.

Finalmente, es importante destacar que el desarrollo de IA generativa debe ser guiado por consideraciones éticas y sociales.

Como mencionamos anteriormente, la capacidad de estas tecnologías para generar contenido engañoso o manipulador puede tener consecuencias graves en la sociedad.

Por lo tanto, es necesario establecer regulaciones y guías éticas para el desarrollo y el uso de la IA generativa, así como para abordar los posibles impactos negativos que puedan surgir.

Clase 13: Storytelling

¿Qué es el Storytelling?

Desde el inicio de los tiempos, el ser humano ha usado las historias para comunicarse con los demás y poder expresar sus pensamientos e ideas.

Por lo tanto el Storytelling, es el arte de contar una historia que genere impacto mediante palabras, imágenes o sonidos. Se busca crear una “atmósfera mágica” a través de nuestro relato.

El StoryTelling, es el proceso de traducir los análisis de datos a términos simples, con el objetivo de influir en una decisión o acción en particular, la cual suele ser de tipo comercial.

Storytelling en la era digital

El storytelling, se ha arraigado sin dudas en diferentes disciplinas. La aparición de Internet, la revolución de las redes sociales y la evolución de las TICs, han posibilitado a los individuos el compartir sus historias en nuevos formatos (como el podcast, el vídeo o el blog) y recibir retroalimentación constante.

Cuando hablamos de Storytelling en la era digital, estamos hablando de la aplicación de los métodos y técnicas de storytelling más tradicionales pero ahora dentro del mundo y contexto digital.

Hoy el término se asocia con diferentes funciones como ser por ejemplo: visualizaciones de datos, infografías, tableros, ciencia de datos, etc.

¿Por qué hacemos visualizaciones?

Para comunicar patrones interesantes. Alentar el descubrimiento y exploración de un conjunto de datos. Detectar patrones en nuestros datos. Entender nuestros modelos.

Conexión emocional con la audiencia: Sin dudas contar una historia permite empatizar con nuestro público objetivo y generar un mayor engagement.

Recordabilidad: Gracias a un relato con impacto, es muy probable que podamos quedar en la mente de nuestra audiencia, aún luego de que quizás haya pasado un largo tiempo.

Confianza: Nuestra historia humaniza nuestro relato, y eso termina impactando en la confianza de nuestro público objetivo.

Engagement: Efectivamente, al empatizar generamos un vínculo con nuestra audiencia que puede tornarse en una relación a largo plazo.

Componentes del Storytelling

Se necesita de al menos los siguientes componentes: información, historia, objetivo, forma visual.

El paso más importante para crear una visualización de impacto es saber y definir qué queremos decir.

También, es imprescindible que nuestra visualización tenga un propósito claro y establecido que nos permita conectar con la audiencia desde una perspectiva humana

Los elementos de un relato son inicio o planteamiento, un nudo, y un final. Se debe responder las preguntas:

WHY/Por que: Debemos empezar definiendo cuál es el objetivo y el propósito de nuestra presentación por ejemplo: persuadir, motivar, explicar, etc.

WHO / quien: Tenemos que pensar a quién va dirigida nuestra presentación. Por ejemplo: nivel estratégico, táctico u operativo.

What: Definimos qué vamos a presentar para que se adapte a los criterios anteriores. Una buena opción es una lluvia de ideas.

Recomendaciones para contar buena historia

Claridad y concisión:

Buscamos resumir el volumen de datos ofreciendo una lectura simplificada, mediante un lenguaje conciso y claro. Además se requiere ser conciso a la hora de transmitir las ideas.

Insights relevantes:

Identificar la información importante del data storytelling es fundamental, igual que determinar el objetivo que perseguimos: queremos inspirar a nuestro público, divertirlo, informar sobre un producto por ejemplo.

Texto e imágenes:

Si pensamos en nuestro buyer persona podremos elegir imágenes de alta carga informativa que les resulten útiles y, a la vez, den soporte a la información de manera atractiva.

Ponernos en la piel del oyente:

Por más que parezca una obviedad muchas veces nos olvidamos de estos aspectos. Necesitamos empatizar y generar engagement. Podemos hacer uso también de técnicas como por ejemplo el Design Thinking Pensamiento de Diseño.

Errores típicos al contar una historia DS

No usar señales visuales en la narración de datos.

Contar la historia sin contexto

Transmitir falta de confianza (lo cual se evidencia en visualizaciones poco elaboradas y resultados sin respaldo)

No conocer bien la audiencia

Hablar sin un propósito claro

Las ideas presentadas no estén interconectadas

Recordar que los algoritmos no toman decisiones

No se prioriza el storytelling en la forma de presentar los resultados

Se plasman varias historias que hacen que la principal pierda su peso

No llevar los datos y resultados obtenidos al plano del mundo real

Clase 14: Tecnologías Aliadas de la IA y Claves de Aplicabilidad

Herramientas aliadas a la IA

APIs

Las APIs (Application Programming Interfaces) son un conjunto de reglas, protocolos y herramientas que permiten que diferentes aplicaciones se comuniquen y compartan datos y funciones. En otras palabras, las APIs actúan como intermediarios entre aplicaciones y permiten que éstas se integren y funcionen juntas de manera más efectiva.

Las APIs son esenciales en el desarrollo de aplicaciones modernas y son una forma de estandarizar la forma en que se intercambian datos y funciones entre aplicaciones.

Bases de Datos NoSQL y BigData Frameworks

Las bases de datos NoSQL, como MongoDB o Cassandra, son herramientas esenciales para cualquier proyecto que requiera almacenar grandes cantidades de datos no estructurados. Estas bases de datos ofrecen una mayor escalabilidad y flexibilidad que las bases de datos SQL tradicionales.

Los big data frameworks, como Apache Hadoop o Apache Spark, facilitan la implementación de proyectos con requisitos de procesamiento de grandes conjuntos de datos. Estas herramientas proporcionan una infraestructura escalable y distribuida para procesar grandes cantidades de datos de manera eficiente.

Cloud Computing

Es una tecnología que permite a los usuarios acceder a recursos informáticos a través de Internet sin tener que tener servidores o equipos físicos en su propia ubicación.

Esta tecnología ofrece una mayor eficiencia y flexibilidad a las empresas y a los usuarios para acceder a aplicaciones y datos desde cualquier lugar y en cualquier momento.

ETLs

Estas herramientas permiten a los usuarios integrar y transformar datos de diferentes fuentes para su uso en modelos de machine learning de una manera más rápida y sencilla. También permiten alcanzar niveles de desarrollo mayores para gente con poca experiencia en programación.

Algunos de los más populares son: alteryx, easymorph, knime, dataiku, datagram.

Softwares de Visualización

Permiten a los científicos de datos visualizar los datos de manera efectiva y crear informes visuales que ayuden a entender los patrones y tendencias en los datos. Además, ofrecen una amplia gama de opciones de análisis, lo que permite a los usuarios explorar y comprender mejor los datos.

Algunos de los más populares son: tableau, powerBI, google Data Studio, qlikView, excel*

Consideraciones clave para aplicar un proyecto tecnológico

Seguridad Informática

Se refiere a la protección de los sistemas informáticos y los datos que se almacenan y procesan en ellos. Esto incluye la prevención de accesos no autorizados, la protección contra virus y otros programas maliciosos, y la prevención de brechas de seguridad que puedan comprometer la información sensible.

La seguridad informática es un factor crítico en cualquier proyecto tecnológico, ya que una falla en la seguridad puede tener consecuencias graves, como la pérdida de datos, la interrupción de los servicios, la exposición de información confidencial o incluso el daño a la reputación de una empresa o organización.

Usabilidad

La usabilidad se refiere a la facilidad con la que un usuario puede interactuar con un sistema o producto tecnológico.

Es importante que los sistemas sean intuitivos y fáciles de usar para los usuarios, ya que esto mejora su experiencia y aumenta la probabilidad de que lo utilicen de manera efectiva.

Interoperabilidad

La interoperabilidad se refiere a la capacidad de diferentes sistemas o productos para trabajar juntos de manera efectiva.

Es importante que los sistemas sean capaces de intercambiar información y trabajar juntos sin problemas, especialmente cuando se integran con otros sistemas o productos existentes en una empresa o organización.

Principio de Parsimonia

El principio de parsimonia o navaja de Occam es un principio filosófico que se utiliza en el mundo tecnológico para tomar decisiones basadas en la simplicidad. Según este principio, si existen varias explicaciones para un mismo problema, se debe elegir la más simple y que requiera menos supuestos o recursos adicionales.

Este principio se aplica en el desarrollo de software y sistemas tecnológicos para evitar la complejidad innecesaria y mantener los sistemas simples y eficientes. De esta manera, se pueden evitar errores y problemas futuros que podrían surgir como resultado de sistemas excesivamente complicados.

Escalabilidad

La escalabilidad es la capacidad de un sistema para crecer y adaptarse a un aumento en la demanda sin comprometer su rendimiento. En otras palabras, un sistema escalable es aquel que puede manejar un mayor volumen de usuarios o procesos sin afectar su funcionamiento.

La escalabilidad es un factor clave en el desarrollo de sistemas tecnológicos, ya que permite a las empresas y organizaciones crecer sin tener que invertir constantemente en nuevos sistemas o recursos. Además, un sistema escalable puede mejorar la experiencia del usuario y aumentar la eficiencia del sistema.

Deuda Tecnológica

La deuda tecnológica es el costo de mantener y actualizar sistemas tecnológicos que han sido diseñados sin tener en cuenta su mantenimiento a largo plazo. En otras palabras, es el precio que se paga por no mantener actualizado el software, hardware y otros recursos tecnológicos en el tiempo.

Esta deuda puede surgir por diversas razones, como una falta de inversión en tecnología, una mala planificación o una falta de capacidad de actualización. Es importante tener en cuenta que, al igual que la deuda financiera, la deuda tecnológica no desaparece y puede generar problemas y costos mayores en el futuro si no se aborda adecuadamente.

Clase 15: Encoders

Encoders

En el contexto del preprocesamiento de datos y el machine learning, un encoder se refiere a una técnica que se utiliza para convertir variables categóricas (también conocidas como variables cualitativas o variables no numéricas) en una forma numérica o binaria que pueda ser utilizada por los algoritmos de machine learning.

Los encoders son necesarios porque muchos algoritmos de machine learning requieren que los datos de entrada sean numéricos. Sin embargo, las variables categóricas representan características o atributos que no se pueden representar directamente como números. Por lo tanto, los encoders se utilizan para asignar valores numéricos o binarios a las categorías, de modo que los algoritmos puedan trabajar con ellas de manera efectiva.

Tipos de Encoders

Existen diferentes tipos de encoders, cada uno con su enfoque único para la codificación de variables categóricas.

Algunos de los encoders más comunes incluyen: One-Hot Encoder, el Label Encoder, Ordinal Encoder, Binary Encoder, Count Encoder, Target Encoder

Cada tipo de encoder tiene sus propias características y se elige en función de las propiedades específicas de los datos y los requisitos del problema.

One-Hot Encoder

Es una técnica comúnmente utilizada para codificar variables categóricas. Convierte cada categoría en una representación binaria de múltiples columnas, donde cada columna

representa una categoría y se asigna un valor binario (1 o 0) según la presencia o ausencia de esa categoría en cada observación.

El One-Hot Encoder es adecuado para variables categóricas sin una relación ordinal clara y cuando el número de categorías únicas es manejable. Sin embargo, puede aumentar significativamente la dimensionalidad del conjunto de datos si hay muchas categorías únicas.

En Python, se puede implementar el One-Hot Encoder utilizando la función `get_dummies()` de Pandas o la clase `OneHotEncoder` del módulo `sklearn.preprocessing`.

Label Encoder

Asigna a cada categoría un número entero único. Cada categoría se reemplaza por su correspondiente número. Esta codificación es útil cuando existe una relación ordinal entre las categorías, es decir, cuando hay un orden inherente en las categorías.

El Label Encoder transforma las categorías en valores numéricos enteros secuenciales, lo que facilita su uso en algoritmos de machine learning. Sin embargo, es importante tener en cuenta que esta codificación puede generar una falsa relación de orden en variables categóricas donde no existe tal orden.

En Python, se puede utilizar la clase `LabelEncoder` del módulo `sklearn.preprocessing` para aplicar la codificación de etiquetas.

Ordinal Encoder

También asigna un número entero único a cada categoría, al igual que el Label Encoder. Sin embargo, a diferencia del Label Encoder, el Ordinal Encoder permite manejar variables categóricas con más de una columna.

El Ordinal Encoder se utiliza cuando tienes variables categóricas ordenadas o jerárquicas, donde el orden o la jerarquía entre las categorías son importantes. Por ejemplo, las tallas de ropa (pequeño, mediano, grande) o las calificaciones (baja, media, alta) son ejemplos de variables donde el orden importa.

En Python, se puede utilizar la clase `OrdinalEncoder` del módulo `sklearn.preprocessing` para aplicar la codificación ordinal. Se puede especificar el orden de las categorías utilizando el argumento `categories` al crear una instancia de `OrdinalEncoder`.

Binary Encoder

El Binary Encoder convierte cada categoría en una representación binaria utilizando una secuencia de bits. Asigna una secuencia de bits para cada categoría, donde cada bit representa la presencia o ausencia de una categoría particular.

El Binary Encoder es útil cuando tienes un gran número de categorías únicas y deseas reducir la dimensionalidad del conjunto de datos en comparación con el One-Hot Encoder. El Binary Encoder generalmente requiere menos columnas que el One-Hot Encoder, pero aún así permite capturar información sobre las relaciones entre las categorías.

En Python, se puede utilizar la clase `BinaryEncoder` del paquete `category_encoders` para aplicar esta codificación.

Count Encoder

Reemplaza cada categoría con el recuento de veces que aparece en el conjunto de datos. Asigna el número de ocurrencias de cada categoría como el valor codificado.

El Count Encoder puede ser útil cuando el recuento de las categorías es informativo y tiene una relación con la variable objetivo. En lugar de asignar números arbitrarios a las categorías, el Count Encoder utiliza la información de frecuencia para codificarlas.

En Python, se puede utilizar la clase `CountEncoder` del paquete `category_encoders` para aplicar esta codificación.

Target Encoder

Utiliza la variable objetivo para codificar las categorías. Reemplaza cada categoría con la media de la variable objetivo correspondiente a esa categoría.

El Target Encoder es útil cuando el target tiene una relación con las categorías y deseas capturar esa relación en la codificación. En lugar de utilizar números arbitrarios o frecuencias, el Target Encoder utiliza información del target para asignar valores a las categorías.

En Python, se puede utilizar la clase `TargetEncoder` del paquete `category_encoders` para aplicar esta codificación. Se debe proporcionar la variable objetivo al ajustar el encoder.

Clase 20: Metodologías de Gestión de Proyectos

Metodologías de Gestión de Proyectos

Una metodología de gestión de proyectos es un conjunto de prácticas, técnicas y herramientas que proporciona una guía para planificar, ejecutar y controlar un proyecto desde su inicio hasta su finalización.

Se utilizan para garantizar que el proyecto se complete dentro del plazo y presupuesto establecidos y que se entregue un producto de calidad.

¿Por qué se utilizan?

Las metodologías de gestión de proyectos se utilizan para reducir los riesgos, aumentar la eficiencia, mejorar la comunicación y garantizar la calidad del proyecto.

También ayudan a los equipos a trabajar de manera más estructurada y organizada para lograr los objetivos del proyecto.

¿De dónde surgieron?

Surgieron en la década de los 50's y los 60's, cuando los proyectos militares y de construcción de infraestructura se hicieron cada vez más complejos y se necesitaba un enfoque más estructurado para gestionarlos.

Desde entonces, han evolucionado y se han adaptado a diferentes tipos de proyectos y necesidades empresariales.

Algunas de las metodologías más comunes incluyen: Waterfall o Tradicional, Lean, CPM, Agile

Metodología Waterfall

También conocida como ciclo de vida secuencial, es un enfoque lineal y secuencial de gestión de proyectos.

Surgió en los años 70 a partir de la industria de la construcción y la ingeniería; propuesta por el ingeniero de software Winston W. Royce.

La metodología se basa en la premisa de que cada fase del proyecto debe completarse antes de pasar a la siguiente.

Las fases son análisis, diseño, implementación, testing y mantenimiento. Cada fase tiene su propio proceso y conjunto de actividades que deben completarse antes de pasar a la siguiente.

Los roles más comunes son:

Gerente de proyecto: lidera el equipo y se asegura de que se cumplan los objetivos

Analista de negocio: identifica las necesidades del cliente y las convierte en requisitos para el proyecto.

Desarrollador: diseña y construye el software.

Probador: verifica que se cumplan los requisitos.

Gerente de calidad: garantiza que se cumplan los estándares de calidad .

Gerente de configuración: administra y controla los cambios.

Usuario: proporciona comentarios y aprueba el software antes de su lanzamiento.

Ventajas de la metodología waterfall

Procesos lineales y secuenciales que permiten una gestión clara y fácil de entender.

Documentación detallada que facilita la trazabilidad y el control del proyecto.

Planificación y estimación de costos más precisa.

Desventajas de la metodología waterfall

Las desventajas de esta metodología incluyen:

Falta de flexibilidad y adaptabilidad a los cambios, ya que cada fase debe completarse antes de avanzar a la siguiente.

Los clientes pueden no tener una idea clara del resultado final hasta que se complete el proyecto, lo que puede provocar problemas de satisfacción del cliente.

¿Cuándo aplicar?

La metodología Waterfall se recomienda para proyectos con requisitos claros y estables, con una clara definición de alcance y objetivos.

Es adecuada para proyectos en los que el cliente o usuario final no necesita involucrarse en cada fase del proyecto.

Es una opción sólida para proyectos con requisitos claros y estables, pero puede no ser la mejor opción para proyectos complejos o en constante cambio.

Metodología Lean

La metodología Lean es un enfoque de gestión de procesos que se centra en la eliminación de desperdicios y la maximización del valor para el cliente.

Surgió en la industria automotriz japonesa en los años 50 y se popularizó en todo el mundo a partir de la década de los 90's.

¿Cuándo aplicar?

La metodología Lean se recomienda para proyectos en los que se busca mejorar la eficiencia y la calidad, reducir costos y tiempos de producción y aumentar la satisfacción del cliente.

Es adecuada para proyectos que tienen procesos repetitivos y que se pueden mejorar mediante la eliminación de desperdicios.

Es una opción sólida para proyectos en los que se busca mejorar la eficiencia y la calidad, pero puede no ser la mejor opción para proyectos complejos o en constante cambio.

Metodología CPM (Critical Path Method)

Es una técnica de gestión de proyectos que se utiliza para planificar y controlar el cronograma de un proyecto.

Surgió en la década de los años 50 en los Estados Unidos y ha sido ampliamente utilizada desde entonces.

El objetivo principal de la metodología CPM es ayudar a los gerentes de proyecto a planificar, programar y controlar proyectos complejos, identificando las tareas críticas que deben completarse en un tiempo específico para que el proyecto se complete en su fecha de finalización prevista.

La metodología CPM se basa en la identificación de las actividades del proyecto y en la creación de una red que muestra la secuencia de estas actividades.

Se utilizan herramientas como el diagrama de Gantt y el análisis de camino crítico para implementar la metodología.

Algunos de los roles más comunes son:

Gerente de proyecto: lidera el equipo y se asegura de que se cumplan los objetivos

Líder de equipo: lidera el equipo y garantiza que se cumplan las tareas asignadas.

Trabajador de línea: realiza las tareas de producción.

Ingeniero de proceso: diseña y mejora los procesos de fabricación.

Ventajas de la metodología de CPM

Identificación del camino crítico, que muestra las actividades más importantes y su impacto en el plazo del proyecto.

Gestión de tiempos, que permite ajustar el cronograma si se presentan retrasos.

Capacidad de planificar y controlar el proyecto de manera más efectiva.

Desventajas

Limitaciones en proyectos complejos con muchas variables y cambios constantes, ya que puede ser difícil identificar y controlar todas las actividades.

Su implementación puede ser compleja y requiere una planificación cuidadosa.

¿Cuándo Aplicar?

La metodología CPM se recomienda para proyectos en los que las actividades están secuencialmente relacionadas y es importante cumplir con un plazo específico.

Es adecuada para proyectos grandes y complejos que requieren una planificación y control rigurosos.

La metodología CPM es una opción sólida para proyectos con actividades secuenciales y un plazo específico, pero puede no ser la mejor opción para proyectos complejos o en constante cambio.

Agile

Las metodologías Agile son un enfoque iterativo e incremental para la gestión de proyectos que se centra en la entrega rápida y continua de valor al cliente.

Surgieron en la década de los 90 como una respuesta a las metodologías tradicionales de gestión de proyectos y han ganado popularidad desde entonces.

Manifiesto Agile

En el año 2001, un grupo de expertos en desarrollo de software se reunieron en una estación de esquí en Utah, Estados Unidos, para discutir formas de mejorar el proceso de desarrollo de software.

La reunión dio lugar al Manifiesto Agile, un documento que establece los valores y principios de las metodologías Agile. El Manifiesto Agile surgió como una respuesta a los problemas del desarrollo de software tradicional, que se caracterizaba por procesos pesados y documentación excesiva.

Valores Agiles

VALOR 01: Valorar a los individuos y las interacciones sobre el proceso y las herramientas.

VALOR 02: Software funcionando por encima de la documentación extensiva.

VALOR 03: Valorar más la colaboración con el cliente que la negociación de un contrato.

VALOR 04: Valorar más la respuesta ante el cambio que seguir un plan.

Principios

Los 12 principios del Manifiesto Ágil son:

1. Nuestra mayor prioridad es satisfacer al cliente mediante la entrega temprana y continua de software con valor.
2. Aceptamos que los requisitos cambien, incluso en etapas tardías del desarrollo. Los procesos ágiles aprovechan el cambio para proporcionar ventaja competitiva al cliente.
3. Entregamos software funcional frecuentemente, entre dos semanas y dos meses, con preferencia al periodo de tiempo más corto posible.
4. Los responsables de negocio y los desarrolladores trabajamos juntos de forma cotidiana durante todo el proyecto.
5. Los proyectos se desarrollan en torno a individuos motivados. Hay que darles el entorno y el apoyo que necesitan, y confiarles la ejecución del trabajo.
6. El método más eficiente y efectivo de comunicar información al equipo de desarrollo y entre sus miembros es la conversación cara a cara.
7. El software funcionando es la medida principal de progreso.
8. Los procesos ágiles promueven el desarrollo sostenible. Los promotores, desarrolladores y usuarios debemos ser capaces de mantener un ritmo constante de forma indefinida.
9. La atención continua a la excelencia técnica y al buen diseño mejora la agilidad.
10. La simplicidad, o el arte de maximizar la cantidad de trabajo no realizado, es esencial.
11. Las mejores arquitecturas, requisitos y diseños emergen de equipos autoorganizados.
12. A intervalos regulares el equipo reflexiona sobre cómo ser más efectivo para a continuación ajustar y perfeccionar su comportamiento en consecuencia.

Estos principios complementan los cuatro valores fundamentales del Manifiesto Ágil y proporcionan una guía más detallada para implementar metodologías ágiles en el desarrollo de software.

Variantes de Agile

Existen diferentes modelos de metodologías Agile estructurados, los cuales se adaptan a diversos tipos de proyectos o empresas. Algunos de estos son:

Scrum, Kanban*, Extreme Programming (XP), Lean*, Crystal, FDD

Cada una tiene su propia estructura y enfoque, pero todas se centran en la entrega rápida y continua de valor al cliente.

Roles más Comunes

Las metodologías Agile incluyen roles específicos que pueden variar dependiendo de la variante aplicada. Algunos de los roles más comunes son:

Product Owner: que es responsable de definir los requisitos del cliente y priorizar el trabajo.

Scrum Master: que facilita el proceso y ayuda al equipo a mejorar.

Equipo de Desarrollo: que es responsable de entregar el trabajo.

Stakeholders: personas o grupos que tienen un interés en el proyecto, incluyendo clientes, usuarios finales, gerentes de proyecto y otros miembros del equipo.

Cliente: la persona o entidad que recibirá el producto final y proporcionará retroalimentación y comentarios sobre el mismo.

Ventajas de la metodología agile

Flexibilidad para adaptarse a los cambios en los requisitos del cliente y las condiciones del mercado.

Fomentan la colaboración entre el equipo y el cliente.

Mejorar la calidad del producto.

Reducir el tiempo de comercialización.

Las desventajas de esta metodología incluyen:

Falta de documentación detallada, que puede dificultar la comunicación entre el equipo y los interesados en el proyecto.

Dependen en gran medida del equipo, lo que puede crear problemas si algún miembro del equipo se retira o no cumple con sus responsabilidades.

¿Cuándo aplicar?

Las metodologías Agile se recomiendan para proyectos complejos y cambiantes en los que es difícil definir todos los requisitos al principio.

Son adecuadas para equipos que valoran la flexibilidad y la colaboración con el cliente y que pueden adaptarse rápidamente a los cambios.