Instituto Tecnológico de Costa Rica **Escuela de Computación** Programa en Ciencias de Datos **Curso: Estadistica** Profesor: Ph. D. Saúl Calderón Ramírez	QUIZ 1 Entrega: Domingo 28 de Abril, a través del TEC digital Debe subir un *pdf* con la respuesta, generado con latex (adjunte los archivos .tex asociados). Valor: 100 pts. Puntos Obtenidos: _____ Nota: _____

Nombre del (la) estudiante: **Marco Ferraro Rodriguez**

Carné: **1 1782 0987**

1. Según la charla del Dr. Alexandre Chiavegatto, responda las siguientes preguntas:

   (a) **(10 puntos)** Qué diferencia hay entre la Inteligencia Artificial clásica y la contemporánea?

   Respuesta

   La inteligencia artificial clásica se basa en algoritmos y técnicas que dependen de reglas predefinidas y lógica formal para tomar decisiones y resolver problemas. Este enfoque es determinista y opera dentro de los límites de las reglas establecidas. Por otro lado, la inteligencia artificial contemporánea, especialmente impulsada por el avance del aprendizaje automático y el aprendizaje profundo, se centra en el desarrollo de sistemas que pueden aprender de los datos y mejorar su rendimiento con el tiempo sin estar explícitamente programados con reglas específicas. Esta forma de inteligencia artificial utiliza grandes volúmenes de datos y algoritmos complejos para identificar patrones, hacer predicciones y tomar decisiones, lo que permite una mayor flexibilidad y adaptabilidad en comparación con los métodos clásicos

   (b) **(10 puntos)** Cuáles son las razones de la irrupción a gran escala del aprendizaje automático?

   Respuesta

   La irrupción del aprendizaje automático a gran escala se debe a una serie de factores interrelacionados que han transformado tanto la capacidad tecnológica como la infraestructura de datos. Primero, la disponibilidad y accesibilidad de grandes volúmenes de datos han sido fundamentales. En un mundo cada vez más digitalizado, la cantidad de datos generados por dispositivos, aplicaciones y plataformas en línea proporciona una rica fuente de información que los algoritmos de aprendizaje automático pueden utilizar para aprender y mejorar.

Además, los avances significativos en hardware han facilitado la ejecución de estos algoritmos de manera más eficiente. La evolución de las GPUs y el desarrollo de hardware específico para tareas de aprendizaje automático, como las TPUs, han reducido drásticamente los tiempos de procesamiento. Como se mencionó en la charla, lo que antes tomaba semanas en procesarse, ahora puede realizarse en cuestión de horas, lo que permite experimentaciones más rápidas y un ciclo de innovación más corto.

La innovación continua en los algoritmos de aprendizaje automático también ha jugado un papel crucial. Con cada avance, estos algoritmos se vuelven más capaces de manejar tareas complejas, desde el procesamiento del lenguaje natural hasta el reconocimiento de imágenes y la toma de decisiones autónoma. Esta mejora constante ha ampliado las aplicaciones prácticas del aprendizaje automático en industrias tan diversas como la salud, la automoción, las finanzas y más allá. Por último, la inversión del mercado en tecnologías emergentes ha sido un motor importante para la adopción del aprendizaje automático. Las startups centradas en la inteligencia artificial y el aprendizaje automático han atraído significativas inversiones de capital, lo que ha permitido una rápida iteración y desarrollo de productos. Este flujo de capital ha fomentado un entorno donde la innovación es rápida y las nuevas aplicaciones son constantemente exploradas y puestas en práctica. En la charla están delimitados por los tópicos de Big Data, Capacidad computacional y avances técnicos.

(c) **(10 puntos)** Cuál es la contribución más importante a las técnicas de aprendizaje automático hasta la fecha?

Respuesta

La contribución más importante a las técnicas de aprendizaje automático hasta la fecha es la capacidad de los algoritmos para aprender de grandes volúmenes de datos y mejorar las decisiones en diversos campos, incluyendo la salud. Esta capacidad se destaca especialmente en el contexto de la inteligencia artificial aplicada al sector salud, donde el aprendizaje automático permite analizar y entender la complejidad de los datos de salud para mejorar las decisiones clínicas y los resultados de los pacientes

(d) **(20 puntos)** Explique e investigue, usando como referencias publiaciones del Dr. Chiavegatto, en qué consiste la primer aplicación mostrada?

Respuesta

Dentro del contexto de la charla, se destaca la aplicación de técnicas de aprendizaje automático (Machine Learning, ML) en la gestión de la pandemia de COVID-19 como una herramienta valiosa para mejorar la atención de la enfermedad y mitigar la propagación del virus. Los modelos predictivos desarrollados mediante ML tienen la capacidad de identificar a pacientes con un alto riesgo de complicaciones graves, lo que permite intervenciones tempranas y medidas de aislamiento más efectivas tanto en entornos hospitalarios como comunitarios.

Una de las primeras aplicaciones presentadas por el Dr. Alexander Chiavegatto se enfoca en el uso de técnicas de aprendizaje automático para predecir el pronóstico negativo de COVID-19 en pacientes hospitalizados. Este enfoque emplea modelos predictivos que analizan datos clínicos y de laboratorio para identificar a los pacientes con mayor

probabilidad de experimentar condiciones críticas, como la necesidad de ingresar a la Unidad de Cuidados Intensivos (UCI), el uso de ventilación mecánica o el fallecimiento.

En el estudio "A multipurpose machine learning approach to predict COVID-19 negative prognosis in São Paulo, Brazil", se analizaron datos de 1040 pacientes diagnosticados con COVID-19 en un hospital de São Paulo, Brasil. Se utilizaron cinco algoritmos de ML para predecir el pronóstico negativo de la enfermedad, incluyendo la necesidad de ingreso en UCI, el uso de ventilación mecánica o el fallecimiento. Los modelos fueron entrenados con una muestra aleatoria del 70% de los pacientes y validados con el 30% restante. Los resultados revelaron un alto rendimiento predictivo, con un AUROC promedio de 0.92, una sensibilidad promedio de 0.92 y una especificidad promedio de 0.82. Las variables más influyentes para la predicción fueron la relación de linfocitos por proteína C-reactiva, la proteína C-reactiva y la Escala de Braden.

Estos hallazgos subrayan la eficacia de los algoritmos de ML en la predicción de desenlaces negativos en pacientes con COVID-19, lo que puede brindar un valioso apoyo a los profesionales de la salud en la toma de decisiones clínicas. Además, la identificación temprana de pacientes con alto riesgo puede facilitar la aplicación de estrategias de aislamiento y tratamiento específicas, reduciendo así la carga sobre los sistemas de salud y limitando la transmisión del virus.

(e) **(20 puntos)** Explique en qué consiste(n) y como funciona(n) las métricas utilizadas para este primer proyecto?

Respuesta

En el estudio "A multipurpose machine learning approach to predict COVID-19 negative prognosis in São Paulo, Brazil", se utilizaron varias métricas para evaluar el rendimiento de los modelos de aprendizaje automático desarrollados para predecir pronósticos negativos de COVID-19 en pacientes hospitalizados. Las métricas principales utilizadas fueron el Área Bajo la Curva de Característica Operativa del Receptor (AUROC), la sensibilidad y la especificidad.

- El AUROC (Área Bajo la Curva de Característica Operativa del Receptor) es una métrica comúnmente utilizada para evaluar el rendimiento de un modelo de clasificación binaria. El AUROC mide la capacidad del modelo para distinguir entre clases (en este caso, pronósticos negativos y no negativos de COVID-19). Un valor de AUROC de 1.0 representa una clasificación perfecta, mientras que un valor de 0.5 indica un rendimiento no mejor que el azar. En el estudio, se reportó un AUROC promedio de 0.92, indicando un alto grado de precisión en la capacidad del modelo para predecir resultados negativos en pacientes con COVID-19.
- La sensibilidad, también conocida como tasa de verdaderos positivos, mide la proporción de positivos reales que fueron identificados correctamente por el modelo como tales. Es decir, indica qué tan bien el modelo puede identificar a los pacientes que realmente tendrán un pronóstico negativo. En este estudio, la sensibilidad promedio fue de 0.92, lo que significa que el modelo fue capaz de identificar correctamente el 92% de los casos que efectivamente resultaron en un pronóstico negativo.
- La especificidad, también conocida como tasa de verdaderos negativos, mide la proporción de negativos reales que fueron identificados correctamente por el modelo

como tales. Esta métrica indica qué tan bien el modelo puede identificar a los pacientes que no tendrán un pronóstico negativo. En el estudio, la especificidad promedio fue de 0.82, indicando que el modelo fue capaz de identificar correctamente el 82% de los casos que no resultaron en un pronóstico negativo.

Estas métricas son fundamentales para evaluar la utilidad clínica de los modelos de aprendizaje automático en contextos médicos, especialmente en situaciones críticas como la pandemia de COVID-19, donde las decisiones rápidas y precisas pueden tener un impacto significativo en los resultados de los pacientes.

(f) **(30 puntos)** Relacione y explique, con los conceptos vistos en clase hasta la fecha, el porqué de la necesidad de aplicar técnicas de *aprendizaje por transferencia* según lo expuesto por el Dr. Chiavegatto en problemas reales donde se usa el aprendizaje automático?

Respuesta

El Dr. Alexander Chiavegatto, en su exposición, destacó la importancia y la necesidad de aplicar técnicas de aprendizaje por transferencia en problemas reales donde se utiliza el aprendizaje automático. Esta necesidad surge de varios factores clave que se relacionan con los desafíos inherentes al aprendizaje automático en contextos prácticos, especialmente en el ámbito de la salud.

- En muchos problemas reales, especialmente en el sector de la salud, existe **una escasez significativa de datos etiquetados**. Esto se debe a que la obtención de etiquetas precisas puede ser costosa, requerir experticia especializada, y en muchos casos, ser logísticamente complicada. El aprendizaje por transferencia permite utilizar modelos preentrenados en grandes conjuntos de datos (generalmente disponibles en dominios relacionados pero más amplios) y afinarlos con el conjunto de datos específico del problema que tiene menos etiquetas. Esto ayuda a superar la barrera de la escasez de datos etiquetados y mejora la precisión del modelo en el dominio específico
- **Entrenar modelos de aprendizaje profundo desde cero requiere una cantidad significativa de recursos computacionales y tiempo**, lo cual puede ser prohibitivo en muchos contextos aplicados. Al utilizar el aprendizaje por transferencia, se aprovecha el conocimiento aprendido por modelos previamente entrenados en tareas similares, lo que reduce considerablemente el tiempo y el costo asociados con el entrenamiento de nuevos modelos desde cero. El aprendizaje por transferencia introduce un conocimiento previo que ayuda a mejorar la generalización del modelo en el nuevo dominio.
- En muchos casos, especialmente en la salud, los problemas pueden variar significativamente entre diferentes poblaciones, condiciones geográficas o características demográficas**. El aprendizaje por transferencia permite adaptar modelos desarrollados para una configuración a nuevas configuraciones con mínimos ajustes**, lo que es más eficiente que desarrollar nuevos modelos desde cero para cada variante del problema.