

Interpretazione di una rete neurale per compliance normativa in ambito bancario

Banca Virtuosa, istituto di riferimento nel settore finanziario, ha identificato l'esigenza di migliorare la trasparenza e la comprensibilità dei modelli di intelligenza artificiale utilizzati nei propri sistemi. Per raggiungere questo obiettivo, Banca Virtuosa ha lanciato un progetto mirato all'implementazione di tecniche di Explainable AI (XAI), in conformità con la normativa vigente sulla trasparenza bancaria.

Attualmente, Banca Virtuosa utilizza modelli di classificazione pre-addestrati per analizzare e classificare dati finanziari critici. Tuttavia, la mancanza di trasparenza nelle decisioni di questi modelli può compromettere la fiducia dei clienti e limitare la capacità della banca di migliorare i propri sistemi in modo mirato. Identificare e correggere gli errori di classificazione è cruciale per garantire accuratezza e affidabilità nei servizi offerti.

Benefici della Soluzione

- Trasparenza nelle Decisioni del Modello:** Implementando tecniche di XAI come Grad-CAM, LIME, SHAP, Integrated Gradients e Occlusion Maps, Banca Virtuosa sarà in grado di generare mappe di salienza che mostrano visivamente quali elementi influenzano le decisioni del modello. Questo incremento di trasparenza migliorerà la fiducia dei clienti e degli stakeholder, dimostrando l'affidabilità e la spiegabilità delle operazioni del sistema di classificazione.
- Miglioramento Continuo delle Performance:** Analizzando le mappe di salienza, Banca Virtuosa potrà identificare con precisione le aree in cui il modello commette errori, sia nelle classificazioni corrette che in quelle errate. Questa analisi dettagliata permetterà di apportare miglioramenti mirati al modello, ottimizzando le sue performance e riducendo il rischio di interpretazioni errate dei dati.
- Conformità Normativa:** Il progetto garantirà che le decisioni dei modelli di intelligenza artificiale siano spiegabili, in linea con i requisiti normativi vigenti. La trasparenza delle decisioni AI è essenziale per la conformità normativa e la governance aziendale, particolarmente in settori regolamentati come quello finanziario.
- Promozione dell'Innovazione:** L'utilizzo di tecniche avanzate di XAI all'interno di Banca Virtuosa promuoverà l'innovazione nel campo dell'intelligenza artificiale. Questo rafforzerà la posizione della banca come pioniere nell'adozione di tecnologie avanzate, consentendo di offrire ai clienti soluzioni sempre più sofisticate e affidabili.

Dettagli del Progetto

- Fase 1: Utilizzo di un Modello di Classificazione Pre-Addestrato**
 - Modello:** Utilizzare un modello pre-addestrato, come DenseNet, dalla libreria torchvision.
 - Dataset:** Applicare il modello a un dataset di immagini, ad esempio MNIST, per esplorare le sue decisioni di classificazione.
- Fase 2: Generazione di Mappe di Salienza**
 - Tecniche di XAI:** Implementare tecniche come Grad-CAM, LIME, SHAP, Integrated Gradients e Occlusion Maps per generare mappe di salienza del modello.

3. Fase 3: Report Finale

- **Descrizione del Dataset:** Dettagliare l'origine, la struttura e le caratteristiche del dataset utilizzato.
- **Analisi delle Mappe di Saliienza:** Confrontare le mappe di salienza per classi corrette ed errate per identificare e comprendere gli errori del modello.
- **Sistema Spiegabile (Opzionale):** Descrivere un sistema completamente spiegabile che potrebbe eseguire la stessa classificazione, offrendo ulteriori insights sulle decisioni del modello.

Obiettivi del Progetto

1. **Comprensione del Modello:** Utilizzare tecniche di XAI per ottenere una comprensione approfondita del funzionamento interno del modello pre-addestrato.
2. **Visualizzazione delle Decisioni:** Visualizzare in modo chiaro e interpretabile quali elementi influenzano le decisioni del modello attraverso le mappe di salienza.
3. **Identificazione degli Errori:** Analizzare le mappe di salienza per identificare e comprendere gli errori del modello, distinguendo tra classificazioni corrette ed errate.
4. **Creazione di Sistemi Spiegabili:** Se possibile, sviluppare o descrivere un sistema completamente spiegabile che possa effettuare la stessa classificazione, fornendo ulteriori insights sulle decisioni del modello.

Motivazione del Progetto Le tecniche di Explainable AI sono essenziali per Banca Virtuosa per migliorare la trasparenza, ottimizzare le performance dei modelli e garantire la conformità normativa. Con questo progetto, la banca mira a rafforzare la fiducia dei clienti, migliorare l'efficienza operativa e promuovere l'innovazione nel campo dell'intelligenza artificiale.

Modalità di consegna:

Link pubblico a notebook di Google Colab