Università degli Studi di Padova

Bug Pharma

E-mail: bugpharma10@gmail.com

Verbale esterno del 05-11-2021

Zucchetti S.p.A. - Login Warrior

Approvazione	-
Redazione	Andrea Salmaso
	Nicla Faccioli
Verifica	-
Stato	Redatto
Uso	Esterno

1 Informazioni generali

1.1 Luogo e data dell'incontro

- **Luogo**: videoconferenza Zoom;
- **Data**: 12-11-2021;
- **Ora di inizio**: 16:00;
- **Ora di fine**: 16:45.

1.2 Presenze

- **Totale presenze**: 7;
- **Presenti**:
  - Lorenzo Piran;
  - Michele Masetto;
  - Silvia Giro;
  - Nicla Faccioli (segretaria);
  - Sara Nanni;
  - Andrea Salmaso;
  - Nicholas Sertori.
- **Assenti**: nessuno
- **Partecipanti esterni**:
  - Gregorio Piccoli (Zucchetti S.p.A.).

2 Domande e risposte

1. **Che tipo di sito serve? C'è altro oltre alla visualizzazione dei dati, come ad esempio il caricamento dei dati da parte dell'utente? Che livello di interazione con il sito serve da parte dell'utente?**

   - Il caricamento dei dati tramite csv dovrebbe stare nell'applicazione.
   - Il tema principale è la visualizzazione:
     (a) con la libreria `D3.js` vengono generati grafici a partire dai dati caricati;
     (b) successivamente si va a discrezione del fornitore.

     Centrale quindi l'interazione con l'utente.

2. **Qual è il punto centrale del progetto? Solo la preparazione e l'analisi dei dati o dobbiamo preoccuparci anche della sicurezza?**

   - Idea principale: utilizzo del prodotto da parte di un sistemista che esegue un monitoraggio dei dati di login a vari applicativi aziendali da parte dei dipendenti.
     (a) I dati di login vengono forniti in un file `CSV`;
     (b) Il file viene caricato nel sistema per l'esplorazione dei dati;
     (c) Il sistema visualizza vari grafici (i quali devono essere scelti dal fornitore).
   - Per quanto riguarda la sicurezza dei dati:
     – Se il file contenente i dati viene caricato in locale: è un tema marginale;
     – Se il file viene caricato nel server per una più facile gestione (data la quantità di dati): è necessario valutarla. Per quanto sensata sia come scelta, dal punto di vista del progetto potrebbe portare via tempo a cose più interessanti (come il Machine Learning).
       Idea: Si potrebbe segnalare al committente, il quale saprà di dover aumentare il budget.

3. **Che tipo di dati ci sono nei file `CSV`? Consigli su come utilizzarli?**

   - È stato aperto un file di esempio e ne sono stati mostrati i punti salienti.
     I dati che l'azienda ci fornirà contengono:
     – Numero utente;
     – Data dell'evento (dato **importante**);
     – Tipo di evento:
       * Login;
       * Errore login;
       * Logout.
     – Applicativo di accesso;
     – IP sorgente;
     – Altri dati non necessari.
   - Conviene dividere la data nelle sue parti (giorno della settimana, mese, settimana dell'anno, ecc) per sfruttarla meglio, così come la fascia oraria (ore e minuti).
   - **Obiettivo finale**: vedere come si distribuiscono.

4. **Quali sono i parametri per giudicare un login sospetto?**

   - Nessun grafico potrà supportare tutte le coordinate contenute nel file `CSV`;
   - Problemi centrali:

  - Scelta delle coordinate da mettere nel grafico;
  - Filtrazione dei dati per ricavarne la massima utilità.
- Sarà necessario valutare il comportamento, a livello prestazionale, del grafico, data la mole di dati che dovranno essere visualizzati.
  Sono stati dati diversi spunti:
  - Visualizzare in sovrimpressione solo i dati di un utente, lasciando sullo sfondo una nebbiolina con i dati degli altri utenti;
  - Utilizzare qualche algoritmo/libreria di ML, fornito/a dall'azienda (es: `PCA`, `t-SNE`, `UMAP`) per riduzione dimensionale (ridurre il numero di coordinate):
    * `PCA`: riduzione lineare;
    * `t-SNE`, `UMAP`: non lineare, preservano le strutture locali a scapito della disposizione generale (esaltazione della distanza).
- Dopo aver ottenuto una visualizzazione ottimale delle informazioni si può considerare l'introduzione di algoritmi di ML che capiscano quello che l'occhio umano riesce a vedere (clusterizzazione, novelty detection, domain approximation), come ad esempio `DBScan`, `HDBScan`, `One-Class SVM`.

5. **Che tipo di formato preferiscono per i manuali da consegnare (utilizzo ed espansione)?**

   È possibile scrivere i manuali in `markdown` e usare il software `Pandoc` per tradurli in formato `pdf`, `LaTeX` o altro.

6. **Consigli generali relativi agli anni precedenti?**

   Attenersi alle tecnologie che sono state indicate come consigliate.