

HOTEL REVIEWS

Estrazione di Insight di Business tramite Sentiment Analysis

DATA ANALYTICS

Anno Accademico 2022-2023

Versione aggiornata al 22/06/2023

A cura di
Nicolò Urbani - 856213
Nada Mohamed - 857606
Alessia Rubini - 851890

Indice

Indice.....	2
Introduzione.....	4
Sentiment Analysis.....	4
Dataset.....	6
Analisi esplorativa.....	8
Preparazione dei dati.....	12
Dataset Preprocessing.....	12
Ridefinizione Score.....	12
Undersampling.....	13
Review data preprocessing.....	14
Word Cloud.....	16
Vettorizzazione.....	18
Modelli di classificazione.....	20
Scelta del modello.....	20
Split dei dati.....	21
Logistic Regression.....	22
Random forest.....	23
Lexicon based.....	24
NRC.....	24
AFINN.....	25
Confronto risultati.....	26
Model explanation.....	27
Feature importance.....	28
Model Explanation - SHAP.....	28
Prediction Explanation - LIME.....	30
Insight.....	33
Dashboard.....	33
Componenti Dashboard.....	33
Struttura Dashboard.....	34
Data Visualization.....	34
Interazione con il dataset.....	36
Explanation.....	39
Scenari.....	42
Hilton Hotel London.....	42
Dataset ChatGPT.....	43
Analisi Finali.....	44
Insights.....	44
Obiettivi Iniziali.....	44

Lunghezza Recensioni.....	45
Limiti.....	45
Conclusioni.....	47

Introduzione

Nell'era digitale odierna, i consumatori si affidano sempre di più alle valutazioni e alle recensioni rilasciate da altri utenti come strumenti decisionali, specialmente quando si tratta di pianificare viaggi verso nuove destinazioni.

In questo contesto, le recensioni e i punteggi (score) assegnati dagli utenti assumono un ruolo di rilievo, in grado di influenzare significativamente la reputazione di un hotel e di impattare in modo indiretto sui relativi guadagni.

L'obiettivo principale di questo progetto è quello di identificare parole che possano fungere da indicatori di recensioni positive o negative, e comprendere come tali parole influenzano il risultato complessivo di una certa recensione.

Attraverso l'applicazione di tecniche avanzate di Sentiment Analysis e Natural Language Processing (NLP), ci si propone di analizzare il contenuto delle recensioni scritte dagli utenti, al fine di comprendere come tali parole impattino sul giudizio complessivo di un hotel.

L'elaborato si estende sia al settore degli affari (business) sia a quello dei consumatori (customer). Per quanto concerne il lato aziendale, l'individuazione delle parole che hanno un maggiore impatto sul giudizio di una recensione permetterà alle strutture alberghiere di identificare quali aspetti del proprio hotel deve migliorare per mantenere un servizio di qualità. Questo consentirà loro di adeguarsi alle aspettative dei clienti e di preservare una reputazione positiva, che si traduce in maggiori guadagni.

D'altro lato, i consumatori potrebbero trarre vantaggio dalle analisi presentate in quanto viene fornito loro uno strumento utile a scoprire quali parole sono più efficaci per influenzare positivamente o negativamente la reputazione di un hotel. In tal modo, potranno scrivere recensioni più incisive e, al contempo, diventeranno più consapevoli rispetto all'impatto di determinate parole contenute nelle recensioni.

Infine, viene proposta una dashboard che effettua sentiment analysis pressoché real time, in modo di fornire uno strumento user-friendly per fare in modo che l'utente (sia business che consumer) possa sfruttare le potenzialità dei modelli proposti per estrarre insight significativi dai dati e prendere decisioni data-driven.

Sentiment Analysis

Al fine di raggiungere gli obiettivi prefissati si applicano varie tecniche di NLP e Sentiment Analysis per costruire modelli in grado di stabilire il sentiment di una recensione, cioè se una recensione è ritenuta positiva o negativa.

La domanda naturale che ci si potrebbe porre è "se ho già uno score rappresentativo per ciascuna recensione, a che serve fare previsione con Sentiment Analysis?".

La Sentiment Analysis è uno strumento che fornisce una maggiore confidenza del giudizio dei consumatori nei confronti delle strutture, in quanto analizza approfonditamente ciascun aspetto descritto nel testo.

Il Reviewer Score potrebbe mostrarsi incoerente per due semplici motivi:

- il recensore può esprimersi complessivamente soddisfatto seppur menzionando aspetti negativi dell'hotel, assegnando uno score sommariamente positivo;
- il recensore può esprimersi complessivamente insoddisfatto seppur menzionando alcuni aspetti positivi dell'hotel, assegnando uno score sommariamente negativo.

Pertanto, si cerca di spingersi oltre il punteggio in sé per focalizzarsi sugli aspetti specifici riportati nel testo di ciascuna recensione, affinando la comprensione (*fine grade understanding*).

Potrebbero, inoltre, emergere comportamenti simili tra utenti o trend che delineano pattern nascosti all'interno del set di dati.

Dataset

Il dataset utilizzato in questo progetto è intitolato “515K Hotel Reviews Data in Europe” ed è reperibile all’indirizzo [Sentiment analysis with hotel reviews | Kaggle](#).

Si tratta di un dataset contenente recensioni di diversi hotel di lusso situati in Europa raccolte dalla piattaforma di prenotazioni online [Booking.com](#) nel periodo compreso tra il 2015 e il 2017.

Il dataset è stato ottenuto mediante *scraping* del sito e tutte le informazioni che contiene sono già di dominio pubblico, nonostante siano di proprietà originaria Booking.

Il contesto dei dati fa riferimento ad un ampio insieme di 515000 recensioni di clienti su ben 1492 hotel di lusso distribuiti in diverse località europee.

Di seguito si riporta un estratto del dataset:

	Hotel_Name	Average_Score	Reviewer_Score	Negative_Review	Positive_Review	Reviewer_Nationality
0	Hotel Arena	7.7	2.9	I am so angry that i made this post available...	Only the park outside of the hotel was beauti...	Russia
1	Hotel Arena	7.7	7.5	No Negative	No real complaints the hotel was great great ...	Ireland
2	Hotel Arena	7.7	7.1	Rooms are nice but for elderly a bit difficul...	Location was good and staff were ok It is cut...	Australia
3	Hotel Arena	7.7	3.8	My room was dirty and I was afraid to walk ba...	Great location in nice surroundings the bar a...	United Kingdom
4	Hotel Arena	7.7	6.7	You When I booked with your company on line y...	Amazing location and building Romantic setting	New Zealand

In particolare, il file CSV del set di dati è composto da 17 campi, di seguito elencati:

- Hotel_Address: indirizzo della struttura
- Review_Date: data in cui il consumatore ha pubblicato la recensione
- Average_Score: punteggio medio dell’hotel, calcolato in base all’ultimo commento pubblicato nell’ultimo anno
- Hotel_Name: nome dell’hotel
- Reviewer_Nationality: nazionalità del cliente che ha pubblicato la recensione
- Negative_Review: recensione negativa data dal revisore sulla struttura. Se il revisore non ha fornito una recensione negativa, il campo viene compilato con “No negative”
- Review_Total_Negative_Word_Counts: numero totale di parole nella recensione negativa
- Positive_Review: recensione negativa data dal revisore sulla struttura. Se il revisore non ha fornito una recensione negativa, il campo viene compilato con “No positive”
- Review_Total_Positive_Word_Counts: numero totale di parole nella recensione negativa
- Review_Score: punteggio assegnato dal cliente all’hotel, basato sulla sua esperienza personale
- Total_Number_of_Reviews_Reviewer_Has_Given: numero di recensioni che il cliente ha fornito in passato
- Total_Number_of_Review: numero totale di recensioni valide ricevute dall’hotel
- Tags: tag assegnati dall’autore della recensione all’hotel
- days_since_review: tempo trascorso tra la data della recensione e la data di acquisizione dei dati

- Additional_Number-of_Scoring: Alcuni ospiti hanno semplicemente assegnato un punteggio al servizio, senza fornire una recensione. Questo numero indica quanti punteggi validi senza recensione sono presenti
- lat: latitudine della struttura
- lng: longitudine della struttura

Da una prima analisi del dataset emerge che il testo delle recensioni disponibili sul dataset è già privo di punteggiatura, ciò sarà utile in fase di data cleaning.

In particolare, le feature ritenute utile per le analisi del progetto sono:

- Hotel_Name
- Reviewer_Nationality
- Negative_Review
- Positive_Review
- Review_Score
- Total_Number_of_Reviews_Reviewer_Has_Given
- Review_Date

Alcune di queste feature vengono utilizzate puramente per analisi esplorativa, altre non utilizzate sono ricavabili, come il numero di parole della recensione.

Al fine della sentiment analysis verranno utilizzate solo le feature:

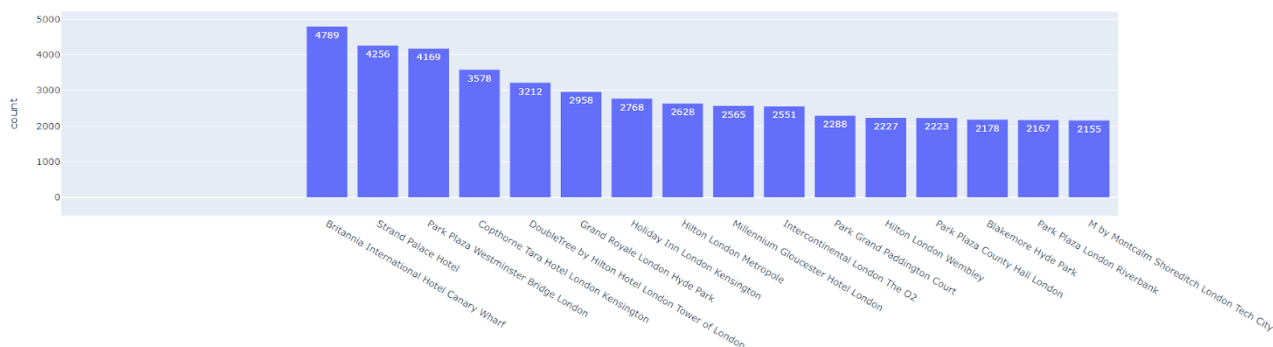
- Reviewer_Nationality
- Negative_Review
- Positive_Review

L' HILTON HOTEL LONDON viene escluso dal dataset utilizzato per le analisi al fine di utilizzarlo nella dashboard al fine di valutare le performance (come visibile nella sezione [Hilton Hotel London](#)).

Analisi esplorativa

L'analisi esplorativa rappresenta una fase essenziale in ambito di analisi dati in quanto consente di ottenere una visione complessiva del set di dati.

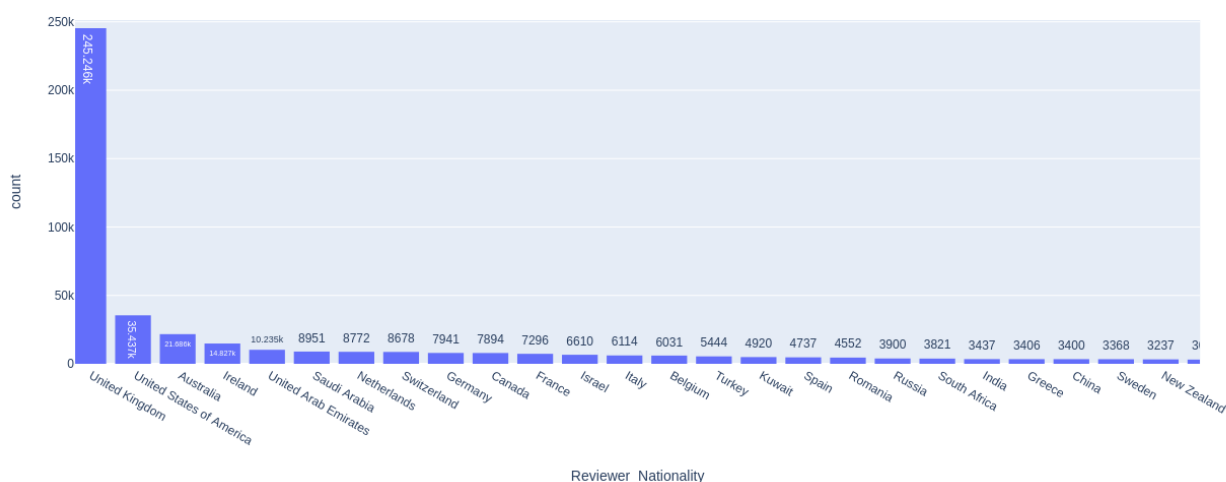
Come riportato dal grafico in basso, tra gli hotel di lusso che collezionano maggior numero di recensioni, ve ne sono diversi situati a Londra.



In realtà quanto descritto non sorprende: tra tutti i siti di prenotazione di hotel online, Booking è quello che dispone della collezione più vasta per quanto concerne la città di Londra, specie in ambito di strutture di lusso.

Come precedentemente accennato, le opinioni raccolte nel dataset sono scritte da clienti provenienti da tutto il mondo, più precisamente, da ben 227 Nazioni differenti.

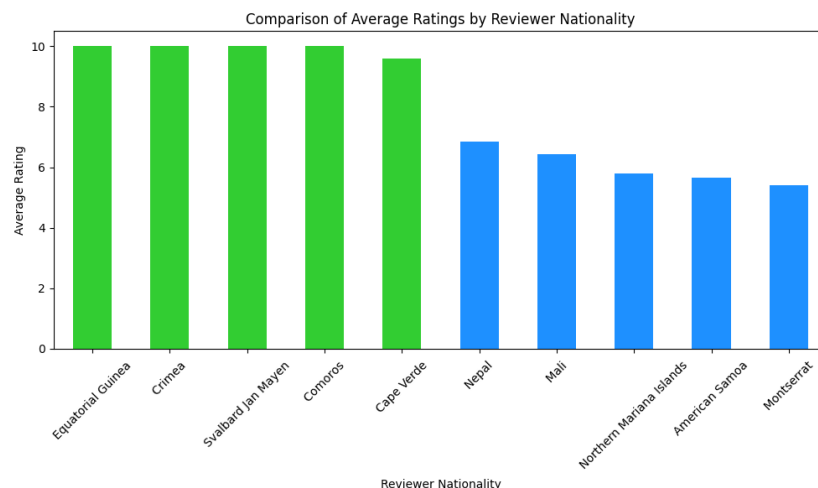
In particolare, col seguente grafico, si vogliono rappresentare i primi 25 Paesi - in termini di quantità di recensioni raccolte - da cui provengono i consumatori.



Emerge subito che i recensori provenienti da UK sono nettamente i più presenti tra tutti, essendo quasi 250k. A seguirli - anche se nettamente con quantità inferiori - vi sono i clienti provenienti dagli USA, poi quelli da Australia, Irlanda ed Emirati Arabi.

Tutte le altre Nazioni riportate contano utenti sempre al di sotto dei 10,000.

L'assegnamento di un certo score da parte di un cliente può essere influenzato dalla sua Nazionalità di provenienza?



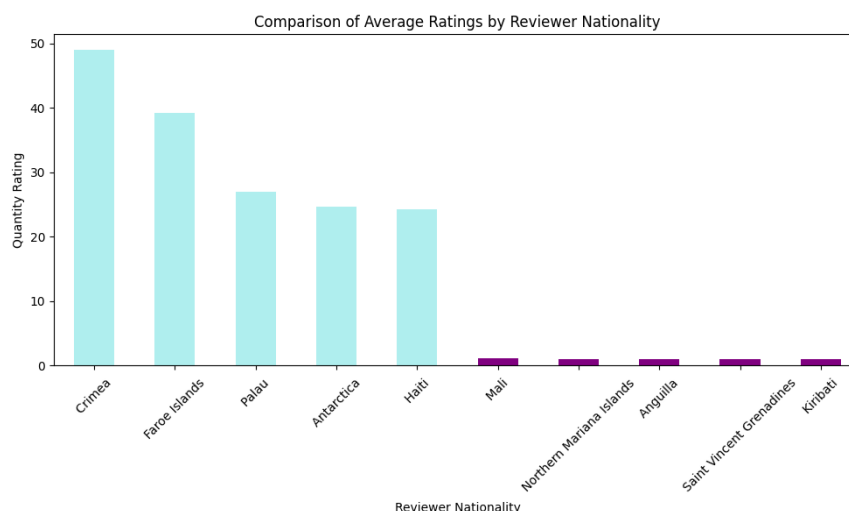
Da quanto evince dal grafico si può affermare che sì, probabilmente il luogo di provenienza del cliente può influire sul voto che assegna alla struttura.

Non a caso le prime cinque Nazioni citate sono località turistiche ricche a loro volta di strutture alberghiere lussuose. Pertanto, clienti di tali nazionalità si potrebbero rivelare più predisposti a giudicare gli hotel di lusso presenti nel dataset con score molto alti.

I restanti cinque Paesi (in blu) sono, invece, quelli che hanno assegnato punteggio più basso; si tratta di Nazioni considerate a reddito medio-basso o in via di sviluppo; quindi potrebbero essere più restii a rilasciare votazioni alte.

Ovviamente quanto detto è frutto di deduzione logica; per rispondere con certezza alla domanda sarebbe più opportuno condurre test statistici specifici, ma non è oggetto del presente elaborato.

Stabiliti quali sono i cinque Paesi più “generosi” e i cinque più “avari” con l’assegnamento dei voti, si vuole ora determinare quali sono le cinque Nazioni da cui derivano il maggior numero di score più alti e le cinque da cui si raccoglie la maggior quantità di punteggi peggiori.

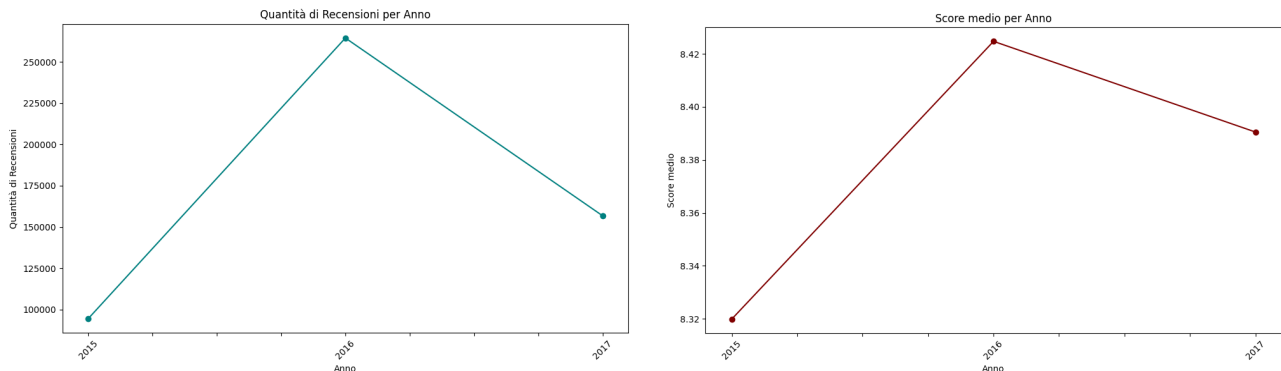


Rispetto al grafo precedente nella “top five” rientra solo la Crimea, mentre nelle cinque peggiori restano sia Mali che Northern Mariana Islands.

Il ch  suggerisce che solo la Crimea ha effettivamente uno score complessivo molto alto su cui pi 

clienti connazionali sono concordi; allo stesso modo, pare che l'assegnamento di punteggi bassi dei consumatori provenienti da Mali e Northern Mariana Island sia un'abitudine comune.

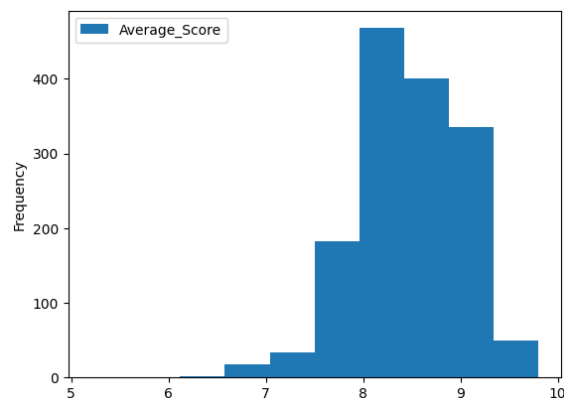
Il set di dati registra recensioni scritte e pubblicate nell'arco di tempo che va dal 2015 al 2017.



I due grafici indicano che l'anno in cui si sono state pubblicate la maggior parte delle recensioni disponibili nel set di dati è quello del 2016; anno in cui si è anche toccato il picco delle medie degli score assegnati, 8.42.

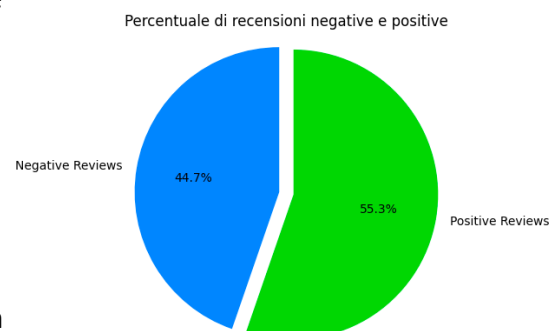
Il punteggio che i recensori possono attribuire a ciascuna struttura si muove in un range che va da 1 a 10.

Come si evince dal diagramma, la maggior parte delle recensioni ha un punteggio superiore a 6.5, tanto che le barre relative ai punteggi più bassi non sono neppure visibili.



Il dataset predispone su due colonne distinte la divisione tra testo di recensioni positive e testo di recensioni negative, come mostrato nel grafico a torta. Confrontando quanto descritto dal grafico a barre precedente con il presente grafico a torta si nota dell'incoerenza.

Infatti, la rappresentazione a destra relativa alla percentuale di recensioni negative e positive, diverse da "No negative", "No positive" e valori nulli, non riporta un gran squilibrio tra le due categorie.



Eppure, il precedente grafico a barre, mostra chiaramente che le recensioni positive sono nettamente maggiori di quelle negative.

Il ch  riassume l'obiettivo su cui verte l'intero progetto: dimostrare che quanto espresso dalle parole durante la stesura di una recensione, non   sempre coerente con il punteggio che gli si associa.

Tale incongruenza pu  costare caro alla reputazione dell'hotel, in quanto il suo score potrebbe influenzare futuri clienti nella scelta di soggiornare o no presso la struttura, e potrebbe non essere in linea con i testi delle recensioni associate.

Preparazione dei dati

Dataset Preprocessing

Al fine di lavorare con una buona qualità dei dati, il dataset è stato preprocessato.

Per prima cosa sono stati eliminati eventuali record duplicati.

Mediante una funzione di conteggio si è osservato che tra le opinioni espresse vi è una forte presenza di recensioni costituite da una o due parole.

Tali giudizi risultano poco significativi al fine delle analisi condotte in questo progetto: da una recensione ci si aspetta una spiegazione, o un excursus della propria esperienza; una parola non si ritiene sufficiente a tale scopo.

Pertanto tutte le recensioni con meno di tre parole sono state eliminate dal dataset.

Si tenga conto che lavorando con un set di dati molto ricco (515K di record) l'eliminazione giustificata di determinate righe, non implica una perdita di dati significativa.

Per i fini ultimi del progetto, tenere in due colonne separate aspetti che il sito ha giudicato positivi e aspetti che ha giudicato negativi di una stessa recensione, non presenta alcuna utilità, in quanto si necessita di informazioni inserite in un contesto complessivo.

Motivo per il quale, è stata inserita all'interno del dataset una nuova colonna, "review", che è il risultato della concatenazione tra le due colonne separate "Negative Review" e "Positive Review".

Qualsiasi recensione nulla all'interno del dataset è stata rimossa.

Inoltre, le frasi "No negative" o "No positive" contenute all'interno delle recensioni sono state sostituite con una stringa vuota "".

Ridefinizione Score

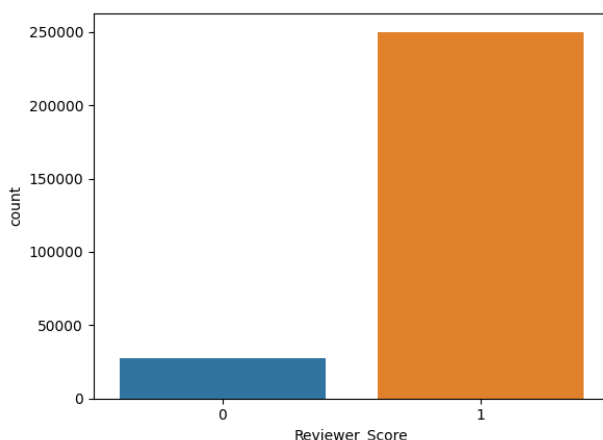
L'obiettivo dell'elaborato è quello di determinare se le parole spese da un utente indicano una recensione positiva o negativa. Quindi si è deciso di ridurre le classi disponibili a due; d'altronde, raramente un cliente pubblica una recensione che esprima sensazioni intermedie. Quello che interessa percepire in questo progetto è la classificazione netta di un giudizio come positivo o negativo.

Per quanto questa decisione implica inevitabilmente una perdita di contenuto, le recensioni neutre non sono utili ai fini delle analisi, e non farebbero che peggiorare le performance dei modelli di classificazione.

Come emerso dalla sezione di analisi esplorativa, vi sono poche recensioni con score molto bassi, quindi si è deciso di suddividere le due classi fissando la soglia 6:

- se il punteggio associato è minore di 6, la recensione viene assegnata alla classe delle recensioni negative 0;
- se il punteggio associato è maggiore o uguale a 6, la recensione viene assegnata alla classe delle recensioni positive 1.

La distribuzione ottenuta delle classi è ora pari a quella mostrata nel grafico:



Si nota subito la complessiva elevata quantità di dati; e, in particolare, le due classi risultano totalmente sbilanciate tra loro: si contano 248580 record per la classe 1 al fronte dei 27186 della classe 0.

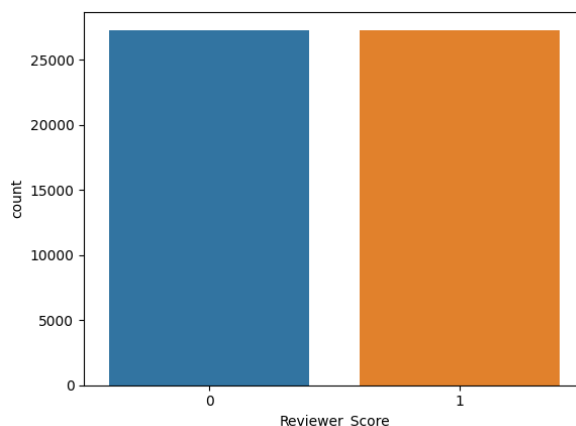
Undersampling

Dunque, il dataset attuale risulta fortemente sbilanciato e le risorse computazionali messe a disposizione dal tool Colab sono limitate non consentendo di effettuare le analisi previste. Risulta necessario intervenire per diminuire la quantità di dati disponibili, e, visto lo squilibrio di distribuzione tra le classi, tale procedura avviene mediante undersampling dell'intero dataset.

Attraverso la tecnica di bilanciamento dati undersampling non verranno creati dati sintetici che potrebbero influenzare i modelli di previsione; e, al contempo, si ottiene un netto ridimensionamento della quantità di dati.

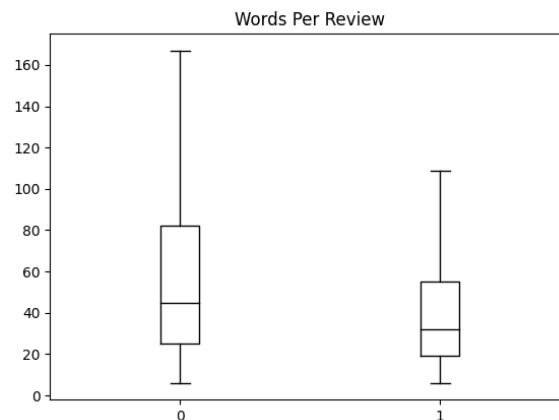
Per effettuare l'undersampling è stata utilizzata la tecnica "RandomUnderSampler" fornita dalla libreria imbalanced-learn. Questa seleziona casualmente un sottoinsieme di campioni della classe maggioritaria - in questo caso la classe positiva - andando ad eliminarle per bilanciare la distribuzione all'interno delle classi.

Così facendo si otterranno le due seguenti classi.



Il che significa che i dati disponibili sono ora circa 60.000: una quantità adatta a lavorare con le risorse computazionali disponibili, e ad ottenere un modello di machine learning che apprenda meglio.

A tal proposito si è condotta una breve analisi “Words per Review” - contando quante parole ci sono mediamente in una recensione - dalla quale emerge un aspetto interessante.



Pare che le recensioni negative siano mediamente più “lunghe”, in termini di quantità di parole contenute al loro interno, rispetto a quelle positive.

Evidentemente, clienti poco soddisfatti che cercano di esaltare tutti gli aspetti negativi emersi dalla propria esperienza, scrivono recensioni più lunghe; mentre, coloro i quali si reputano soddisfatti della struttura in questione, esprimono in modo più conciso gli aspetti positivi che hanno riscontrato.

Review data preprocessing

Nella fase di preprocessing delle singole recensioni, il focus è quello di elaborare i testi al fine di prepararli per essere successivamente elaborati; in altre parole, si entra nel campo del NLP.

Con il termine NLP - acronimo di “Natural Language Processing” (elaborazione del linguaggio naturale) - ci si riferisce a quel campo dell'intelligenza artificiale che si occupa di comprendere, analizzare e generare il linguaggio umano in modo automatico. L'obiettivo del NLP è consentire alle macchine di interagire con il linguaggio naturale in un modo che sia significativo e, soprattutto, comprensibile.

Alcune delle principali aree di studio all'interno del NLP includono procedure che interessano questa fase specifica di preprocessing.

Tra le più importanti vi sono la tokenizzazione, il POS tagging, la lemmatizzazione, lo stemming e la rimozione di stop word.

La tokenizzazione è il processo di suddivisione di un testo in unità più piccole chiamate, appunto, token. Nel caso di studio, il testo di ciascuna recensione viene suddiviso in parole, o meglio dire token.

Il POS tagging, o part-of-speech tagging, è quel processo in cui si assegna una categoria grammaticale (parte del discorso) a ciascuna parola del testo. Questa procedura determina se una certa parola è un verbo, un pronome, un aggettivo, un nome, ecc.; e risulta particolarmente utile per comprendere meglio la struttura grammaticale di un testo.

La lemmatizzazione e lo stemming sono due processi utili a ridurre le parole alle loro forme base o radici comuni.

La prima assegna ad una parola il suo lemma, nonché la sua forma canonica o di dizionario.

Lo stemming, d'altra parte, riduce una forma alla sua radice rimuovendo affissi o terminazioni. Questo non tiene conto del contesto e può produrre radici che non sono necessariamente classificabili come parole reali.

Entrambe le procedure sono utili per la normalizzazione delle parole, riducendo le variazioni grammaticali e semplificando l'analisi del testo. Tuttavia, la lemmatizzazione tende a fornire risultati più accurati; dunque, per questo progetto si è preferito sottoporre i singoli token al processo di lemmatizzazione e non a quello di stemming.

Le stop word sono parole molto comuni, come articoli, preposizioni o congiunzioni, che spesso non portano ad alcun significato informativo per l'analisi del testo. Dunque, la loro rimozione è una tecnica comune nell'elaborazione di un testo in quanto riduce il rumore e migliora l'efficienza dei modelli di analisi linguistica, come nel caso della Sentiment Analysis.

Per il NLP ci si è serviti del tool NLTK. Si tratta di una piattaforma leader per lavorare con dati di linguaggio umano.

L'omonima libreria Python fornisce interfacce semplici da utilizzare per oltre 50 corpora e risorse lessicali come WordNet, insieme ad una collezione di librerie di elaborazione del testo per classificazione, tokenizzazione, stemming, e molto altro.

In particolare, per la rimozione di stop word, fornisce delle liste predefinite per diverse lingue, inclusa l'inglese.

Di seguito, si riportano tutte le modifiche apportate ai testi:

- I caratteri sono stati convertiti a caratteri in minuscolo;
- Gli spazi presenti prima e dopo il testo sono stati eliminati;
- Mediante la funzione "word_tokenize" (della libreria NLTK), il testo è stato suddiviso in singoli token, più immediati da analizzare;
- I caratteri numerici sono stati rimossi;
- Impostando l'insieme di stop words NLTK "english" come lista, è stato possibile inserirvi altre parole considerate non influenti sulle analisi condotte. Si è deciso, però, di escludere - e quindi mantenere all'interno dei testi - le parole "not", "nor" e "against" da tale lista, in quanto si è pensato potessero modificare il significato di alcune frasi. Ottenuto l'insieme finale di stop words, si è passato all'eliminazione delle stesse per ogni recensione disponibile;
- Eventuali token vuoti sono stati rimossi;
- Si è applicato il POS tagging tramite una funzione personalizzata: ogni token è stato classificato come aggettivo, verbo, nome o avverbio
- Si è eseguita la lemmatizzazione del testo mediante comando fornito da NLTK;

Al termine di tutte queste modifiche, una *join* ha permesso di "ricompattare" i token in testi, ottenendo così tutte le recensioni pre processate.

Si riporta un esempio di una recensione iniziale, e della stessa pre processata:

It was a little dark in lobby
dining room but this wasn t really
a problem just hard to see Amazing
location right at the heart of
shopping and tourist sights Staff
spoke excellent English and were
very helpful

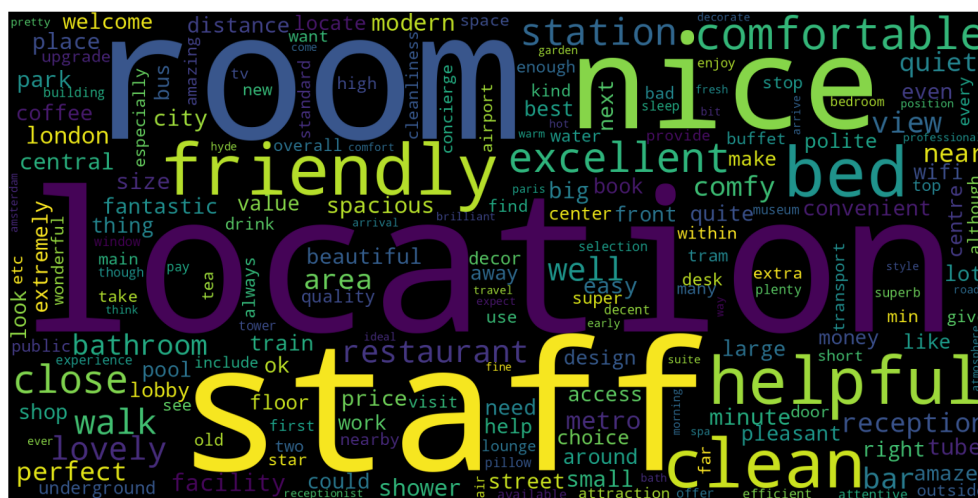
dark lobby dining room problem
hard see amaze location right
heart shop tourist sight staff
speak excellent english helpful

Word Cloud

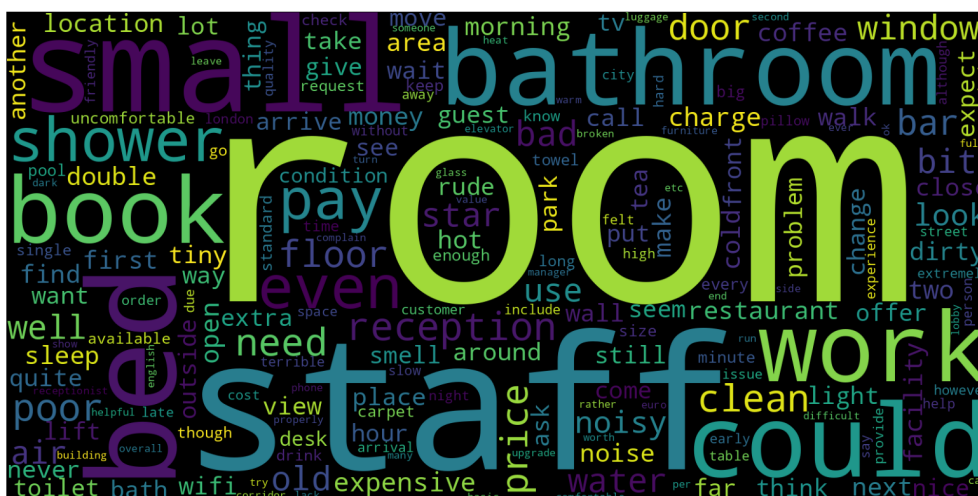
Una volta pre processato il testo di ciascuna recensione è possibile iniziare ad estrarre i primi insights. La Word Cloud, infatti, consente di esplorare visivamente - attraverso dei plot - le parole più frequenti utilizzate dai consumatori per recensire gli hotel di lusso presenti nel dataset.

A tal proposito, si è applicata la procedura di preprocessing sulle colonne originali del dataset che dividevano la recensione in parte positiva e parte negativa, in modo da rappresentare quali sono i vocaboli più frequenti in un caso e poi nell'altro.

La colonna relativa alle recensioni positive ha fruttato il seguente plot:



Quella relativa alle recensioni negative, invece, ha prodotto

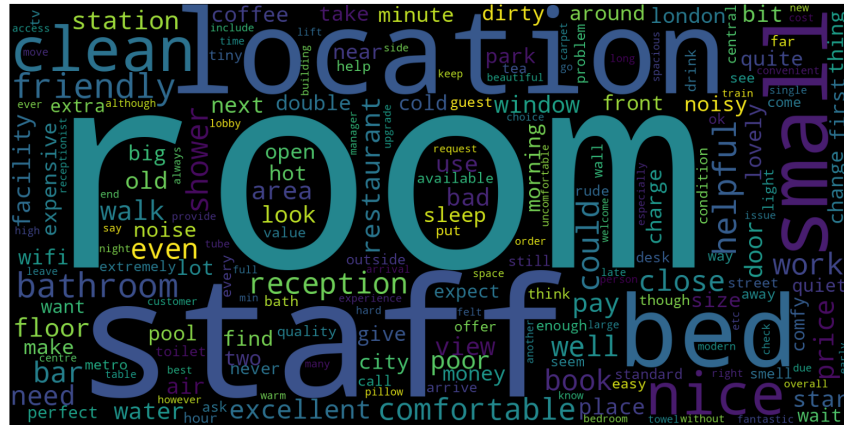


Si nota subito che vi sono degli aspetti che, in base al contesto, cambiano di significato; di fatti, parole come “room” o “staff” sono molto frequenti sia in recensioni negative che positive. Al contempo, vi sono parole che si contraddistinguono per rappresentare aspetti positivi - come "excellent" o “friendly” - o aspetti negativi - come "dirty" o “small”.

In generale, tra i vocaboli più frequenti vi sono sia quelli comuni ad entrambe le classi (*room, location, staff, bed*) che quelli appartenenti ad una sola delle categorie, come: *nice, comfortable,*

friendly.

A supporto di quanto detto si riporta il plot relativo alle word cloud sulla colonna *review* che contiene tutte le recensioni, includendo sia aspetti positivi che negativi.



Lo strumento di Word Cloud consente di farsi un’idea immediata su quelli che sono - a prima vista - i punti di forza e di debolezza degli hotel considerati, fornendo una panoramica visiva delle opinioni degli ospiti.

Vettorizzazione

L'analisi del linguaggio naturale ci consente di passare da una forma di dato non strutturato (testo) a forma di dato strutturato.

NLP (Natural Language Processing) è il campo dell'intelligenza artificiale che si occupa dell'elaborazione automatica e dell'interpretazione del linguaggio naturale da parte dei computer. Consente ai computer di comprendere, analizzare e generare il linguaggio umano in modo significativo.

Tra i diversi approcci disponibili si è scelto di utilizzare quello relativo alla basic vectorization. In particolare, l'obiettivo era quello di riflettere l'importanza di una determinata parola rispetto alle altre presenti nel documento e nel corpus; ragion per cui è stata implementata la rappresentazione del testo mediante **TD-IDF**.

TF-IDF (Term Frequency-Inverse Document Frequency) è una misura statistica che riflette l'importanza di una parola in un documento all'interno di una raccolta (corpus). Calcola un peso per ogni parola in base alla sua frequenza nel documento (TF) e alla sua rarità nel corpus (IDF). Le parole che appaiono frequentemente in un documento ma raramente nel corpus ricevono un peso maggiore.

La TF (term frequency) misura la frequenza di un termine o di una parola in un determinato documento. Mentre, l>IDF (inverse document frequency) misura la rarità del termine in un corpus.

Per le analisi di questo progetto si è deciso di mantenere solo parole con almeno 100 occorrenze, in quanto è un numero non troppo alto - evitando il rischio di introdurre rumore e confusione nei modelli - che permette comunque di ottenere buone performance. Tale scelta viene presa anche in relazione alle risorse computazionali disponibili al fine di lavorare con un numero di feature minore e rendere possibile allenare e testare i modelli.

Utilizzando TF-IDF si ottiene una vettorizzazione context-free, in quanto non tiene conto del contesto, ciò comunque sarà utile in una fase successiva per procedere con l'exploration dei modelli.

Durante la fase di vettorizzazione si lavora solo sulla colonna "review", complessiva di tutte le recensioni pre-processate, ottenendo complessivamente 1505 feature - corrispondenti a 1505 vocaboli.

Di seguito, si riporta la tabella di vettorizzazione relativa ad alcune parole.

	word_able	word_absolutely	word_ac	word_not	word_accept	word_acceptable	word_access
40012	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
44562	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
46997	0.000000	0.000000	0.057590	0.000000	0.000000	0.000000	0.000000
41957	0.000000	0.057084	0.000000	0.000000	0.000000	0.000000	0.000000
41939	0.000000	0.000000	0.000000	0.000000	0.000000	0.070118	0.000000

Come si può osservare dalla tabella riportata, le colonne rappresentano la vettorizzazione di alcune

parole all'interno di 5 recensioni estratte in modo random dal dataset.

I numeri contenuti nella tabella rappresentano i valori di TF-IDF associati a ciascuna parola in base al documento in cui si trovano.

I vettori numerici ottenuti rappresentano le caratteristiche del testo e possono essere utilizzati come input per algoritmi di apprendimento automatico.

Pertanto ora è possibile addestrare modelli di classificazione utili ad etichettare le recensioni come positive o negative in base ai vettori rappresentanti le singole recensioni.

Il set di dati ottenuto, che verrà poi utilizzato per le previsioni conterrà:

- 1505 feature, ognuna rappresentante una parola diversa (risultato della procedura TF-IDF)
- La colonna contenente il Reviewer Score che assume valori 0 o 1 (Negativo - Positivo)

Modelli di classificazione

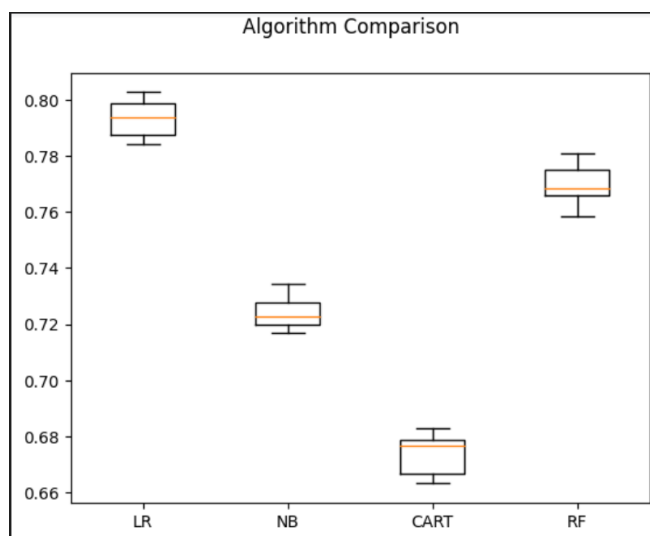
Scelta del modello

Una volta eseguita la vettorizzazione delle recensioni - al fine di rappresentarle come feature numeriche, rendendole adatte all'elaborazione da parte dei modelli di classificazione - è stata utilizzata la tecnica della 10-fold cross validation per valutare quale modello di classificazione è il più promettente fra:

- Random Forest
- Logistic Regression
- Naive Bayes
- CART

La cross validation è una tecnica utilizzata per stimare le prestazioni di un modello in modo più robusto ed evitare problemi di overfitting. Nel caso di studio, il dataset è stato diviso in 10 parti uguali, note come "fold". Successivamente, sono stati svolti l'addestramento e la valutazione dei modelli su ciascuna combinazione delle fold, utilizzandone nove per addestrare il modello e il restante per testare le relative prestazioni. Questo processo viene ripetuto per tutte le combinazioni delle fold, in modo che ciascuna di esse fosse utilizzata almeno una volta come set di test.

L' **accuracy** ottenuta con la 10-fold cross validation applicata ai quattro modelli viene riportata di seguito:



L'accuratezza più elevata è stata ottenuta dalla Logistic Regression (LR) con il valore di 0.79, seguita dal Random Forest (RF) con un'accuratezza media di 0.77. Il modello Naive Bayes (NB) ha ottenuto un'accuratezza media di 0.72, mentre CART ha mostrato un'accuratezza media di 0.67.

Sulla base di questi risultati, è possibile considerare la Logistic Regression (LR) e il Random Forest (RF) come i modelli che hanno ottenuto le prestazioni migliori in termini di accuratezza per la classificazione delle recensioni.

In particolare notiamo che le Random Forest ottengono un intervallo di confidenza leggermente minore rispetto alle Random Forest, presentando un accuracy minore ma una confidenza maggiore rispetto alle performance del modello.

Sono necessarie ulteriori valutazioni al fine di scegliere il modello più adeguato al caso in analisi, pertanto si procede analizzando nel dettaglio i modelli più promettenti.

Split dei dati

Una volta selezionati i modelli che si adattano meglio al nostro dataset, si sono suddivisi i dati in set di addestramento e di set di previsione. Questo ha permesso di valutare le prestazioni dei modelli sui dati di test, allenandoli sui dati di train.



Per effettuare questa suddivisione, è stata usata la funzione `train_test_split` della libreria `scikit-learn`.

Dovendo confrontare i risultati tra diversi modelli è stato impostato il parametro `random_state` a 1234 per garantire la riproducibilità dei risultati.

Lo split viene effettuato, come di default in `scikit-learn`, in modalità stratified mantenendo la proporzione fra le etichette nei due set ottenuti (train e test).

Logistic Regression

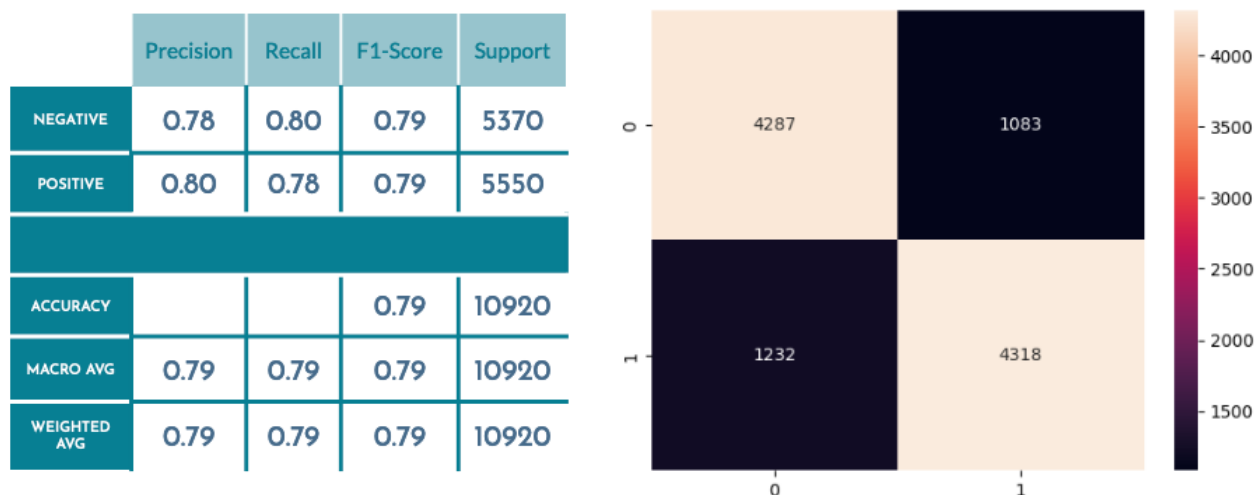
Si analizza la Logistic Regression, in quanto, oltre ad essere il modello migliore emerso dall'analisi iniziale tramite cross validation, è largamente utilizzato nella classificazione binaria.

Il modello viene allenato utilizzando gli iperparametri di default:

```
log_model = LogisticRegression(max_iter = 250, random_state=42)
```

Il tuning degli iperparametri combinato alla cross validation utili ad individuare con quale valore dei parametri si ottengono le migliori performance non si è potuto eseguire a causa di una mancanza di risorse computazionali.

Di seguito si riporta la tabella dei risultati delle metriche di performance e la relativa matrice di confusione:



- **Precision:** rappresenta la percentuale di istanze correttamente identificate come positive (o negative) rispetto al totale delle istanze classificate come positive (o negative).
Nel caso della classe NEGATIVE, abbiamo una precisione di 0.78, il che indica che il 78% delle recensioni classificate come negative dal modello è effettivamente negativo.
Per la classe POSITIVE, abbiamo una precisione del 0.80, che indica che l'80% delle recensioni classificate come positive dal modello è effettivamente positivo.
- **Recall:** rappresenta la percentuale di istanze positive (o negative) correttamente identificate dal modello rispetto al totale delle istanze positive (o negative) presenti nel dataset.
Per la classe NEGATIVE, abbiamo un recall di 0.80, il che significa che il 80% delle recensioni negative presenti nel dataset è stato correttamente identificato come tale dal modello.
Per la classe POSITIVE, abbiamo un recall di 0.78, il che indica che il 78% delle recensioni positive presenti nel dataset è stato correttamente identificato come tale dal modello.
- **F1-Score:** è una media armonica tra precision e recall e fornisce una misura bilanciata delle prestazioni del modello.

Per entrambe le classi, NEGATIVE e POSITIVE, l’F1-Score è 0.79, indicando un buon equilibrio tra precisione e recall.

- **Support:** indica il numero totale di istanze per ciascuna classe nel dataset.
Per la classe NEGATIVE, abbiamo 5370 istanze, mentre per la classe POSITIVE abbiamo 5550 istanze.
- **L'Accuracy** complessiva del modello è 0.79, che rappresenta la percentuale di predizioni corrette rispetto al totale delle predizioni effettuate sul dataset di test.

In generale, i risultati della Regressione Logistica mostrano un buon equilibrio tra precisione e recall per entrambe le classi. L'accuratezza complessiva del modello è inoltre ragionevolmente alta.

Random forest

L'altro modello che è stato selezionato per valutare le relative performance è il secondo in ordine di miglior accuracy ottenuto dall'analisi iniziale, ovvero Random Forest.

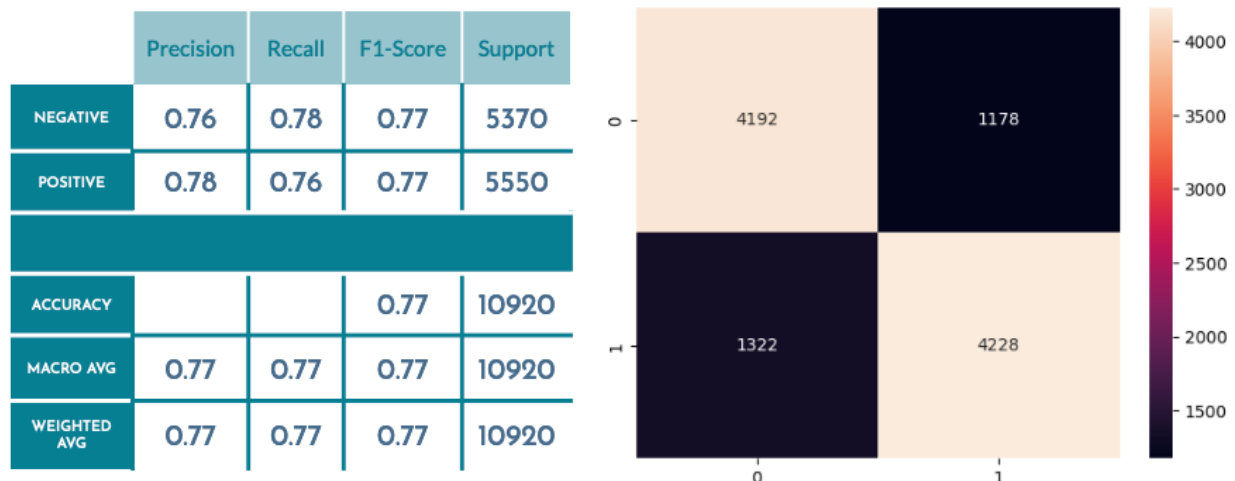
Si sceglie di utilizzare questo modello in quanto, permette di gestire grandi quantità di dati ad elevata dimensionalità e consente di ottenere facilmente la feature importance; infatti, i tool che verranno utilizzati per l'explanation presentano dei tools specifici per approfondire l'apprendimento delle Random Forest.

Il modello viene allenato utilizzando gli iperparametri di default:

```
rf = RandomForestClassifier(n_estimators = 500, random_state = 42,  
criterion='gini')
```

Per le stesse ragioni precedentemente riportate, non è stato possibile effettuare il tuning degli iperparametri combinato alla cross validation.

Si riportano di seguito le performance e la matrice di confusione ottenute applicando il modello di previsione sul test-set, dopo averlo allenato sul train.



- **Precision:** per la classe NEGATIVE è 0.76, il che significa che il 76% delle recensioni classificate come negative dal modello è effettivamente negativo.
Per la classe POSITIVE, abbiamo una precisione del 0.78, che indica che il 78% delle

recensioni classificate come positive dal modello è effettivamente positivo.

- **Recall:** per la classe NEGATIVE è 0.78, il che indica che il 78% delle recensioni negative presenti nel dataset è stato correttamente identificato come tale dal modello. Per la classe POSITIVE, abbiamo un recall di 0.76, il che indica che il 76% delle recensioni positive presenti nel dataset è stato correttamente identificato come tale dal modello.
- **F1-Score** per entrambe le classi, NEGATIVE e POSITIVE, è 0.77, indicando un buon equilibrio tra precisione e recall.
- **Support** indica il numero totale di istanze per ciascuna classe nel dataset. Per la classe NEGATIVE, abbiamo 5370 istanze, mentre per la classe POSITIVE abbiamo 5550 istanze.
- **L'Accuracy** complessiva del modello è 0.77, che rappresenta la percentuale di predizioni corrette rispetto al totale delle predizioni effettuate sul dataset di test.

In generale, i risultati del Random Forest mostrano un buon equilibrio tra precisione e recall per entrambe le classi.

L'accuratezza complessiva del modello è anche ragionevolmente alta, anche se leggermente inferiore rispetto alla Regressione Logistica.

Lexicon based

Al fine di confrontare tecniche diverse per effettuare sentiment analysis, dopo aver utilizzato dei modelli di classificazione, al fine di comparare le performance, si utilizzano due metodi basati sul lessico, il modello NRC e il modello AFINN, per effettuare previsioni sulla positività o negatività delle recensioni.

Questi modelli si basano sull'analisi delle parole presenti nelle recensioni e utilizzano un lessico predefinito per attribuire un punteggio di positività o negatività a ciascuna parola.

Nel contesto di questo studio, i modelli sono stati applicati sia alle recensioni prima che dopo il preprocessing.

L'obiettivo di questa analisi consiste nel valutare la variazione dell'accuratezza dei modelli basati sul lessico prima e dopo l'applicazione del preprocessing.

Ciò ha permesso di esaminare l'effetto del preprocessing sulle capacità di classificazione dei modelli, al fine di predire correttamente la positività o negatività delle recensioni.

NRC

Tale modello utilizza valori binari per indicare se la parola considerata appartiene a una delle **8 emozioni principali**: rabbia, paura, attesa, fiducia, sorpresa, tristezza, gioia e disgusto.

I principali vantaggi del modello NRC sono:

- **Categorizzazione del sentiment:** si etichettano le parole con categorie specifiche di sentiment, quali rabbia, paura, attesa, fiducia, sorpresa, tristezza, gioia e disgusto.
- **Lingue multiple:** sono comprese parole in diverse lingue

Come spiegato in precedenza sono state effettuate due previsioni, una usando le recensioni originali e una usando i dati preprocessati.

Di seguito si riportano le due matrici di confusione e i relativi valori di accuratezza.

Dataset originale		
	0	1
0	6443	16106
1	2058	21633

Accuratezza: 60,7%.

Dataset pre-processato		
	0	1
0	8965	13078
1	3305	19196

Accuratezza: 63,2%.

Dai risultati ottenuti, è possibile osservare che l'accuratezza del modello NRC è migliorata dopo l'applicazione del preprocessing delle recensioni. Prima del preprocessing, l'accuratezza era del 60,7%, mentre dopo il preprocessing è salita al 63,2%.

La matrice di confusione mostra come il modello abbia classificato le recensioni in base alla loro positività o negatività. Dopo il preprocessing, il numero di falsi positivi è diminuito, indicando una migliore capacità del modello di identificare correttamente le recensioni negative.

Allo stesso tempo, il numero di falsi negativi è aumentato, evidenziando un peggioramento nella capacità di individuare le recensioni positive.

Complessivamente questi risultati suggeriscono che il preprocessing delle recensioni ha avuto un impatto positivo sulle prestazioni del modello lexicon-based NRC.

AFINN

Tale modello utilizza valori compresi tra +5 e -5 per determinare quanto un recensione è positiva o negativa.

I principali vantaggi del modello AFINN sono:

- **Semplicità di utilizzo:** assegna un punteggio di sentiment alle parole senza la necessità di considerare categorie specifiche di sentiment
- **Buona copertura delle parole comuni**

Di seguito si riportano le due matrici di confusione e i relativi valori di accuratezza, prima e dopo il preprocessing delle recensioni.

Dataset originale		
	0	1
0	9104	15080
1	1898	23624

Accuratezza: 65,8,7%.

Dataset pre-processato		
	0	1
0	10455	12428
1	2539	21516

Accuratezza: 68,1%.

Dai risultati ottenuti, è possibile osservare che anche l'accuratezza del modello AFINN è migliorata dopo l'applicazione del preprocessing delle recensioni. Prima del preprocessing, l'accuratezza era del 65,8%, mentre dopo il preprocessing è salita al 68,1%.

La matrice di confusione mostra che dopo il preprocessing, il numero di falsi positivi è diminuito, indicando una migliore capacità del modello di identificare correttamente le recensioni negative. Allo stesso tempo, il numero di falsi negativi è aumentato, evidenziando un peggioramento nella capacità di individuare le recensioni positive.

Si può notare dunque un comportamento analogo al modello precedente, in quanto anche in questo caso i risultati suggeriscono che il preprocessing delle recensioni ha avuto un impatto positivo sulle prestazioni del modello, consentendo di ottenere una maggiore accuratezza.

Confronto risultati

Facendo riferimento alle metriche di valutazione come l'accuratezza, relative ai modelli lexicon based, possiamo osservare che il modello AFINN ha ottenuto un valore superiore rispetto al modello NRC sia prima che dopo l'applicazione del preprocessing delle recensioni. Questo indica una maggiore precisione generale del modello nel predire correttamente la positività o negatività delle recensioni.

Il fatto che il modello AFINN classifichi meglio la classe positiva è un risultato importante, poiché le recensioni positive possono essere particolarmente rilevanti per valutare l'opinione degli utenti su un Hotel. Tuttavia, il significativo miglioramento nella classificazione della classe negativa è altrettanto rilevante, poiché ciò indica una maggiore precisione nel riconoscimento delle recensioni negative, le quali possono influenzare gli insight che verranno riportati, per esempio al proprietario dell'Hotel, per indicare gli aspetti da migliorare all'interno della propria struttura.

Complessivamente si può osservare che i modelli lexicon based risultano meno accurati rispetto a quelli ai modelli di Machine Learning.

Si è quindi deciso di procedere utilizzando tecniche basate sul Supervised Classification.

Nonostante la regressione logistica mostri un leggero miglioramento in termini di prestazioni (Accuracy 0.79), abbiamo ritenuto opportuno optare per l'utilizzo dell'algoritmo random forest

(accuracy 0.77). Questa scelta è motivata dalla sua maggiore adattabilità all'explanation del modello, approfondita e commentata nei prossimi capitoli.

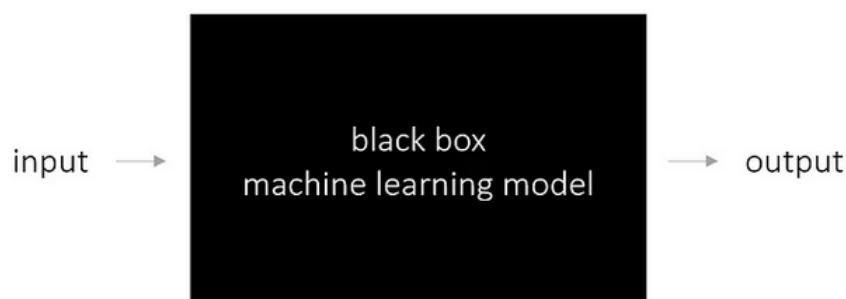
Model explanation

In questa sezione dell'elaborato si effettua l'Explanation - o spiegazione del modello - nonché una pratica utile a comprendere in modo più approfondito i modelli di apprendimento e ad individuare i fattori chiave che influenzano le loro previsioni.

Questo serve a comprendere le scelte dei modelli e analizzare le parole chiave che determinano la classificazione di una relazione come positiva o negativa; il che torna utile rispetto agli obiettivi iniziali.

In particolare, si è effettuata la Model Explanation su *Random Forest*: in quanto modello derivato dagli alberi di decisione, è particolarmente indicato per l'*explainability*.

I modelli di apprendimento automatico sono incredibilmente potenti e in grado di fornire previsioni e risultati accurati in base ai dati di input. Tuttavia, spesso questi modelli sono considerati come "scatole nere" perché la loro logica interna non è facilmente interpretabile dagli esseri umani.



La differenza nell'interpretazione tra gli esseri umani e le macchine è un aspetto importante da considerare. Mentre le macchine possono elaborare grandi quantità di dati e utilizzare algoritmi complessi per effettuare previsioni, gli esseri umani hanno bisogno di spiegazioni comprensibili e intuitive per comprendere appieno le ragioni dietro le previsioni dei modelli di apprendimento automatico.

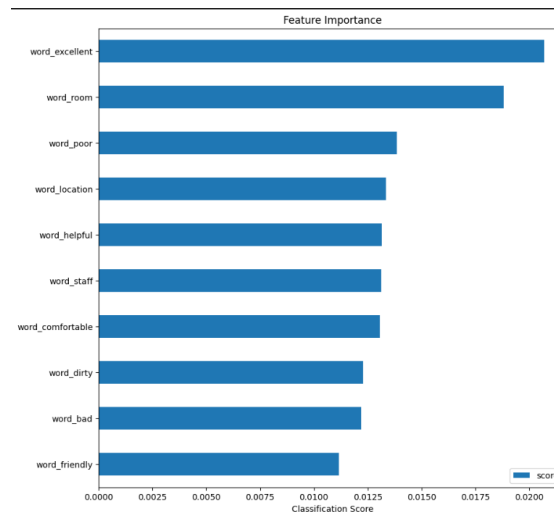
È qui che entrano in gioco gli strumenti di interpretabilità. Comprendere i metodi di interpretabilità utilizzati è fondamentale per evitare di basare le proprie interpretazioni e decisioni su presupposti errati o spiegazioni fuorvianti. Valutare l'importanza delle caratteristiche utilizzate dal modello per effettuare previsioni aiuta a identificare quali aspetti influenzano maggiormente le previsioni stesse.

Questo livello di comprensione dei metodi di interpretabilità contribuisce a una maggiore fiducia e trasparenza nel processo decisionale automatizzato.

Evitando interpretazioni errate o incomplete delle previsioni del modello, è possibile prendere decisioni più informate e mitigare potenziali limitazioni o bias che potrebbero essere presenti.

Feature importance

Il training delle random forest produce la *feature importance*, di seguito riportata nell'istogramma. La feature importance valuta quanto ciascuna feature del set di dati contribuisce alle previsioni complessive del modello. Questa misura può essere ottenuta calcolando la riduzione di impurità media o il guadagno di informazione ottenuti dalla variabile di input in tutti gli alberi della random forest.



Attraverso la feature importance, possiamo identificare le parole che hanno il maggior impatto sulle previsioni del modello.

Nell'algoritmo Random Forest, l'importanza delle caratteristiche viene calcolata valutando l'effetto di ciascuna variabile sul miglioramento della precisione delle previsioni attraverso la riduzione dell'impurità.

L'analisi effettuata calcola l'importanza delle feature a livello locale, nel contesto specifico del modello di Random Forest.

L'importanza delle feature viene calcolata per il modello specifico addestrato utilizzando i dati disponibili e i parametri del modello.

La feature importance può fornire anche un'indicazione dell'importanza delle feature a livello globale, cioè non tiene conto di come le feature siano rilevanti in classi specifiche, cioè non mostra quali sono le feature più rilevanti per l'identificazione di una classe.

Lavorando con una classe binaria (recensione positiva vs recensione negativa) può risultare interessante andare a considerare separatamente le caratteristiche più rilevanti per ciascuna categoria, positiva e negativa. A tale scopo, si procede utilizzando modelli più avanzati.

Model Explanation - SHAP

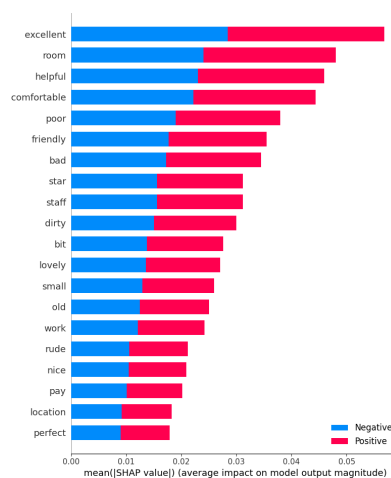
SHAP (Shapley Additive exPlanations) è un tool di Model Explanation ampiamente utilizzato nell'apprendimento automatico. È basato sul concetto di valutazione Shapley della teoria dei

giochi, che assegna un valore di importanza a ciascuna feature di input considerando tutte le possibili combinazioni di feature.

In questa fase del progetto, il framework SHAP, è stato utilizzato per approfondire la precedente analisi di *feature importance* al fine di effettuare un'analisi per classe. Difatti, SHAP fornisce un tool specifico *SHAP tree explainer* utile per l'explination sulle random forest.

SHAP attribuisce un valore (SHAP Value) di contribuzione a ciascuna caratteristica (in questo caso parola), che rappresenta il grado di influenza di quella determinata feature sulla previsione per quella specifica classe.

Di seguito si riporta il grafico ottenuto applicando il framework a quanto ottenuto dalla precedente analisi sul trainset.



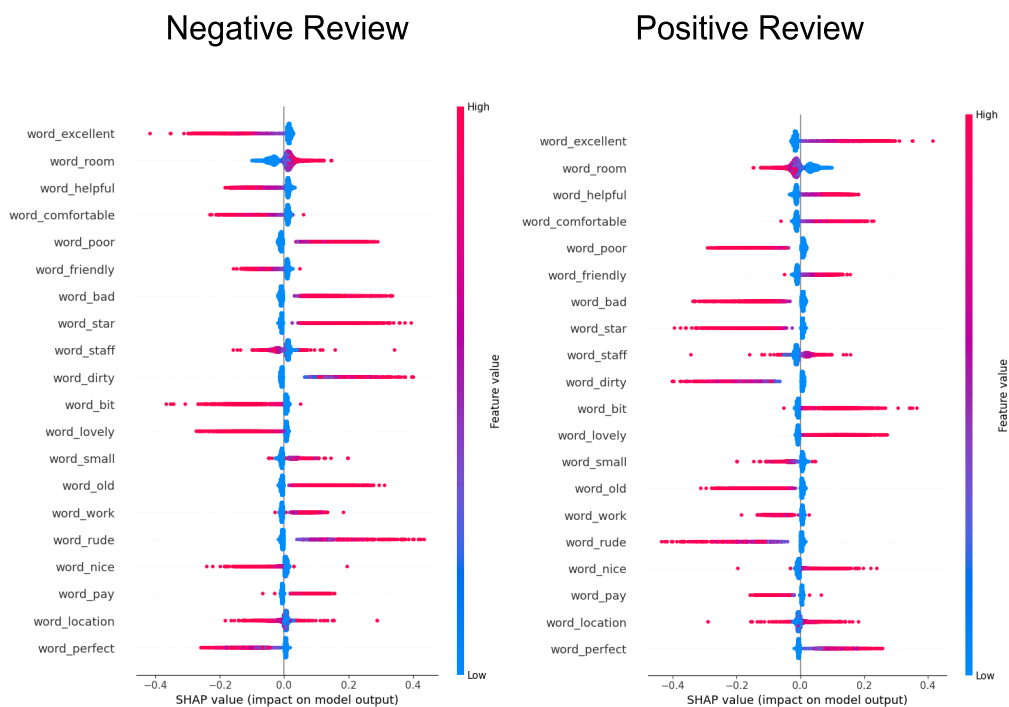
Gli assi del grafico riportato sono:

- asse Y: feature utilizzate nel modello;
- asse X: rappresenta il valore dello Shapley associato ad ogni feature. Lo *Shapley value* è l'indice di misura dell'importanza di una determinata feature. I valori più alti indicano una maggior rilevanza della feature per la predizione del modello.

Tra i vocaboli di maggior rilevanza appaiono *excellent*, *room*, *helpful*, *comfortable*.

Si noti come lo *SHAP Values* risulti all'incirca ben distribuito tra classe positiva e classe negativa.

Pertanto, si procede con un'analisi più dettagliata visualizzando lo SHAP Values per ognuna delle istanze e per ognuna delle feature, indicato dai punti presenti nel grafico di seguito:



In particolare, si ottiene che quasi tutte le feature ottengono SHAP values speculari su entrambe le classi.

Un valore SHAP positivo indica un contributo positivo della feature alla classe positiva, mentre un valore SHAP negativo indica un contributo negativo della feature alla classe positiva.

Invece, per la classe negativa, un valore SHAP positivo indica un contributo negativo della feature, mentre un valore SHAP negativo indica un contributo positivo della feature.

Da ciò ne deriva che le parole positive impattano positivamente sulla classe positiva, e negativamente sulla classe negativa. Quindi nel caso in cui una recensione negativa contenesse la parola 'excellent' ciò andrà a ridurre lo SHAP value, cioè riduce la negatività.

L'analisi svolta permette di concludere che vi sono alcune parole che impattano maggiormente nella classificazione delle recensioni come positive o negative, ma non vi sono parole che determinano una classificazione netta in classe positiva o negativa.

Infatti, la sentiment complessiva sarà determinata da un insieme di parole negative o positive che combinate determinano il sentiment complessivo. Se una recensione contiene parole ad alto impatto come 'excellent', se non conterrà molte parole negative che la contrastano, verrà classificata con maggior probabilità come positiva.

Prediction Explanation - LIME

L'analisi precedente ha consentito di interpretare il modello proposto, le Random Forest, individuando le parole che impattano maggiormente nella classificazione.

Considerato il contesto, ogni recensione avrà aspetti peculiari caratterizzati dal linguaggio umano

spesso con una semantica complessa: si ritiene necessario analizzare le singole recensioni tramite Explanation Locale sulle singole istanze.

Un importante strumento di interpretabilità locale è LIME, che rappresenta l'acronimo di "Local Interpretable Model-agnostic Explanations".

LIME è una tecnica che consente di approssimare qualsiasi modello di apprendimento automatico a scatola nera con un modello locale e interpretabile.

In questo modo, è possibile spiegare ogni singola previsione del modello in modo comprensibile agli esseri umani.

L'interpretabilità dell'IA offerta da LIME rivela ciò che accade all'interno di questi sistemi complessi e aiuta a identificare potenziali problemi come la fuga di informazioni, le distorsioni del modello, la robustezza e la causalità. L'utilizzo di LIME fornisce un framework generico per scoprire le "scatole nere" e fornire una spiegazione del "perché" dietro le previsioni o le raccomandazioni generate dall'IA.

Una caratteristica chiave di LIME è che è "model-agnostic", il che significa che può essere utilizzato con qualsiasi modello di apprendimento automatico, indipendentemente dalla sua complessità o dal metodo utilizzato durante l'addestramento. Questo lo rende uno strumento estremamente flessibile e applicabile a una vasta gamma di problemi di regressione e classificazione.

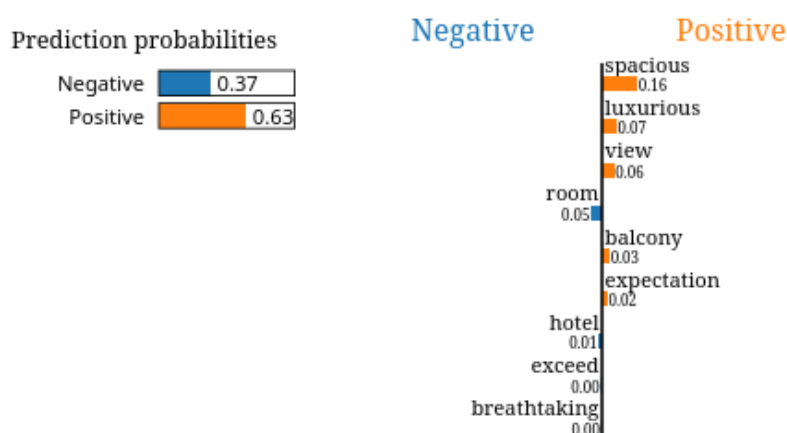
LIME utilizza modelli intrinsecamente interpretabili, nel nostro caso alberi decisionali, per fornire spiegazioni visive comprensibili agli utenti non tecnici.

Questo approccio consente agli utenti di comprendere meglio le ragioni e i fattori che influenzano le previsioni del modello, consentendo loro di prendere decisioni più informate e di fidarsi delle raccomandazioni o delle previsioni generate dall'IA.

Lime viene utilizzato per interpretare la classificazione delle singole recensioni fornite in input per cercare di capire perché sono state classificate in una determinata classe. Di seguito si riportano due esempi:

☐ **Recensione Positiva - Reviewer Score 10**

"This hotel exceeded my expectations. The room was spacious and luxurious, and the view from the balcony was breathtaking."



Tale recensione ha un sentiment positivo, infatti viene rilevato sia dal modello di classificazione e viene spiegato da LIME. Il framework fornisce un livello di confidenza secondo cui la recensione viene classificata positivamente, con il 63% di confidenza la recensione è positiva.

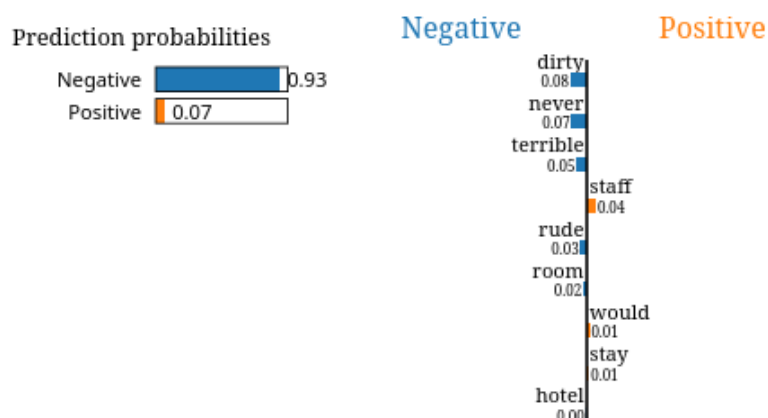
Inoltre, vengono mostrate le parole che hanno impattato maggiormente ossia 'spacious' e 'luxurious'.

Un Impatto leggermente negativo viene dato da 'room'.

Ciò consente all'utente di capire se la recensione inserita ha un impatto positivo o negativo e capire quali aspetti hanno influito maggiormente.

☐ **Recensione Negativa - Reviewer Score 1**

'The hotel was terrible. The rooms were dirty and the staff was rude. I would never stay here again.'



In questo caso la recensione ha un sentiment chiaramente negativo: l'utente si esprime contrariato su molti aspetti dell'hotel.

La recensione viene classificata negativamente con il 93% di confidenza.

Gli aspetti che impattano maggiormente sono la pulizia, avendo la parola 'dirty', la parole 'terrible', tutte parole negative.

La parola 'staff' influisce leggermente positivamente ma viene contrastata da tutte le parole negative presenti, Quindi complessivamente otteniamo un sentiment negativo.

L'exploration locale effettuata con LIME consente di spiegare recensione per recensione; ciò risulta utile sia alla struttura alberghiera per analizzare le singole recensioni ed estrarre gli aspetti positivi e negativi sottolineati, sia per il recensore che riesce ad avere un feedback rispetto a quanto scritto e utilizzare le parole corrette per esprimere un'opinione corretta riguardo alla struttura in analisi.

Insight

Le analisi svolte precedentemente al fine di identificare il modello più performante ed adatto ai casi in analisi, seguite dall'explanation forniscono una serie di risultati utili per estrarre insight utili nel processo decisionale di una struttura alberghiera, per esempio per scegliere quali servizi potenziare oppure quali servizi migliorare o rimuovere in quanto non apprezzati dai clienti.

Dashboard

Al fine di utilizzare i risultati di analisi ottenuti e fornire uno strumento pratico in grado di fornire supporto nelle decisioni viene proposta una Dashboard. La dashboard, sviluppata tramite framework DASH, viene pensata come uno strumento da fornire ad un'ipotetica struttura alberghiera per analizzare le recensioni ottenute.

La dashboard utilizza un dataset di input, in formato CSV contenente i seguenti campi:

- Reviewer Score fornito dall'utente con valori continui da 1-10
- Nazionalità del recensore espressa tramite stringa
- Recensione testuale dell'hotel

Si suppone tale input in una funzione di uso reale venga estratto automaticamente, tramite API oppure tramite web-scraping (se possibile) dalla piattaforma Booking.com oppure da altre piattaforme di review per la struttura in analisi.

La dashboard proposta è eseguibile con qualsiasi dataset avente la struttura predefinita.

Nel nostro caso vengono simulati due scenari di uso della dashboard al fine di testare in modo più reale possibile le performance dei modelli.

Si ipotizzano due scenari al fine di testare lo strumento sviluppato:

- **SCENARIO 1:** Dashboard con dataset contenente le recensioni dell' *Hilton Hotel London* (estratte dal dataset originale *Hotel Reviews 500K*)
- **SCENARIO 2:** Dataset di Recensioni generate casualmente tramite ChatGPT (approccio sperimentale)

In entrambi i casi la dash è equivalente e cambiano solamente i dati utilizzati. La dash utilizza i modelli costruiti nelle analisi precedenti in una situazione ipotetica che potrebbe essere reale.

Componenti Dashboard

In particolare la dash proposta utilizza i seguenti componenti realizzati precedentemente che vengono integrati tramite l'utilizzo del framework Python Dash:

- Modello Random Forest allenato precedentemente ed esportato nel file `Hotel_Review_RF_model.pkl`
- *Vettorizzatore* `tfidf_vectorizer.pkl` allenato precedentemente sul dataset di training
- Funzione di preprocessing `clean_text(text)` che presa in input una recensione applica il preprocessing

Tali componenti vengono integrati nella applicazione che tramite il layout definito e le callback andranno ad utilizzare i componenti per realizzare le analisi richieste.

In particolare l'applicazione è composta da:

- Definizione del layout

- Callback e relative funzioni

Fra le funzioni rilevanti implementate abbiamo la funzione

`sentimentAnalysis(review_processed_array)` che si occupa dato un dataset in input di effettuare il preprocessing del dataset, la vettorizzazione e infine applicare il modello importato per effettuare la sentiment Analysis.

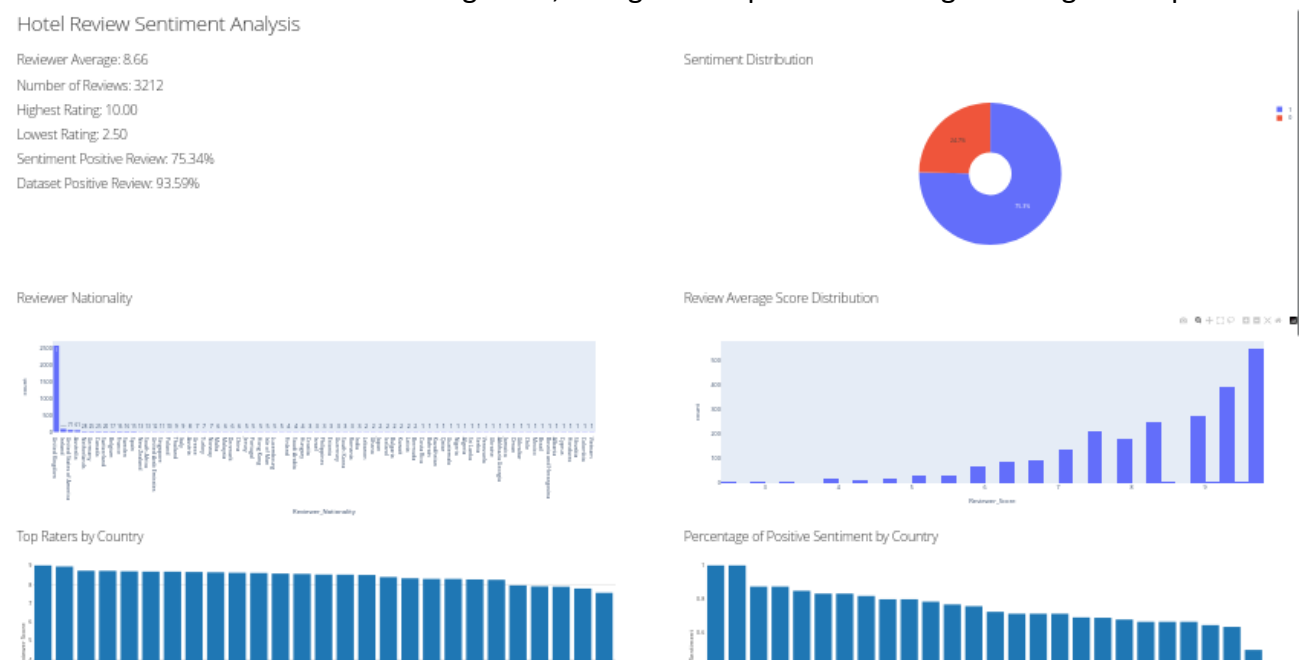
Al termine verrà aggiunta al dataset un colonna che conterrà il risultato di sentiment.

Tutte le componenti definite consentono di presentare i dati tramite analisi descrittive e inoltre consentono di effettuare varie operazioni sul dataset che consentono di rendere la vista interattiva modificando la visualizzazione in base ai dati.

Struttura Dashboard

Di seguito si illustrano le varie componenti che costituiscono la dashboard. Per comodità si riporta la dashboard prodotto nello scenario 1, nonostante le considerazioni fatte siano valide per ogni possibile dataset utilizzato.

La struttura della dashboard è la seguente, di seguito si riportano i dettagli dei singoli componenti:



Data Visualization

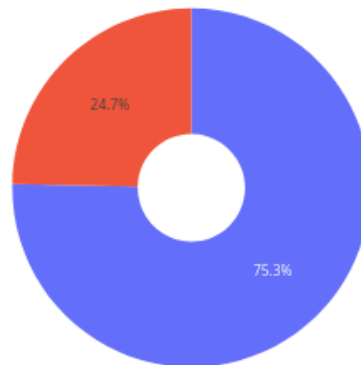
Nella parte iniziale della dashboard vengono indicate alcune analisi riassuntive del dataset, ma anche alcuni insights provenienti dalla sentiment analysis applicata a tutto il dataset di input:

- Reviewer Average: media delle recensioni ricevuto per l'Hotel in analisi
- Number of Reviews: numero di recensioni ricevute (fornisce informazione rilevanza statistica per il campione analizzato)
- Highest Rating: rating più alto ottenuto in tutte le recensioni
- Lowest Rating: rating minore ottenuto in tutte le recensioni
- Sentiment Positive Review: percentuale di recensioni positive ottenute tramite sentiment (es 75% indica la percentuale rispetto a tutte le recensioni di recensioni positive (valore 1) ottenute nella sentiment)

- Dataset Positive Review: numero di recensioni con score ≥ 6 in tutto il dataset rispetto alla recensioni totale espresso in percentuale (es 75% significa che nel dataset il 75% delle recensioni ha uno score ≥ 6)

Viene riportato un grafico a torta dei risultati ottenuti dalla sentiment ciò permette di intuire quale è l'opinione del mio hotel tramite sentiment analysis. In questo caso il 75% delle recensioni è positivo (classe 1) mentre il 25% negativo.

Sentiment Distribution

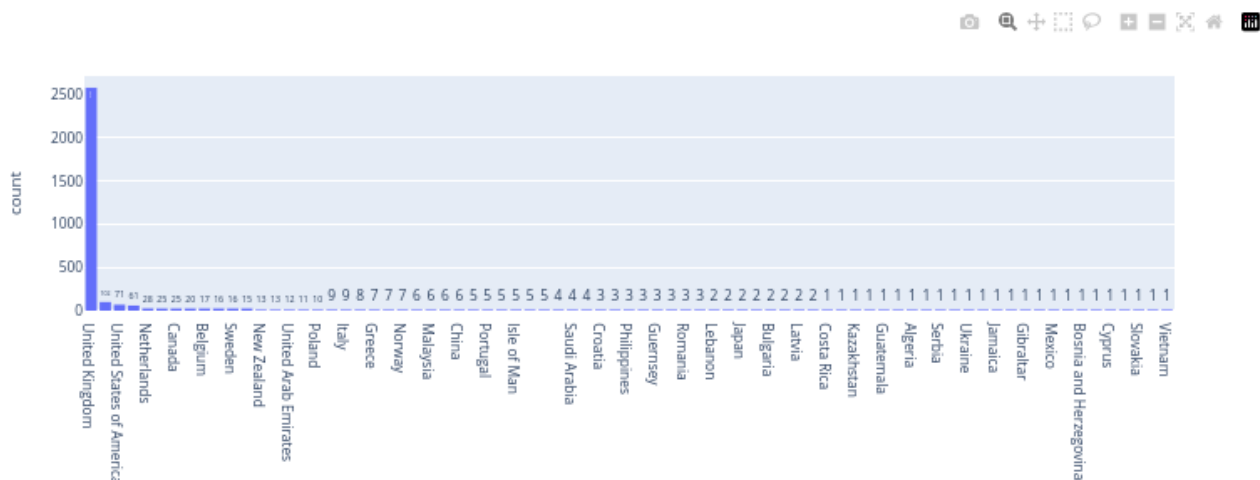


Si riportano analisi descrittive al fine di avere informazione relativa alla clientela presente nel mio hotel in base alla nazionalità al fine di capire il tipo di clientela del mio hotel in analisi e la distribuzione degli score ottenuti.

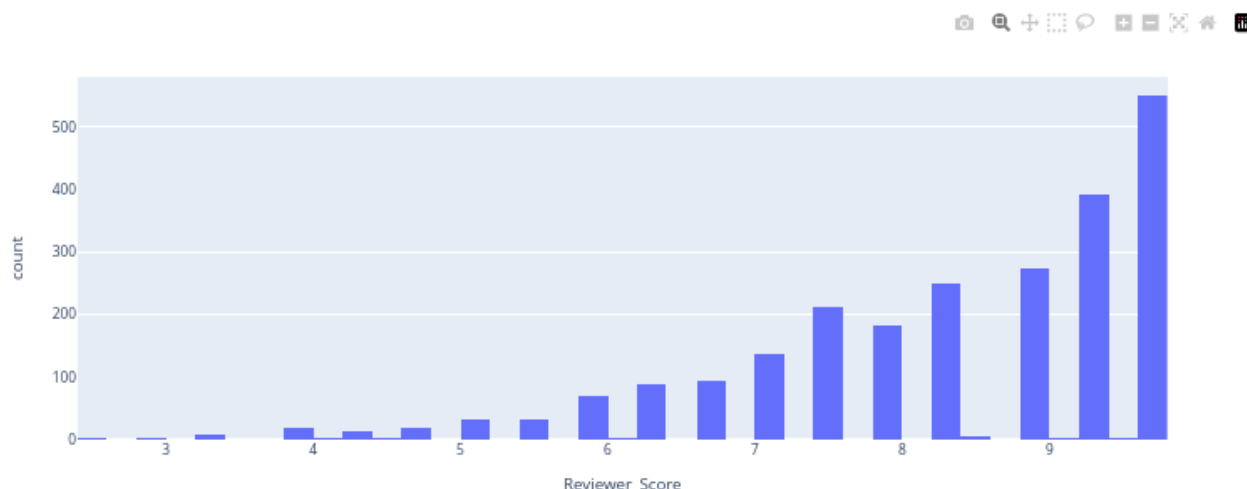
Tale grafico consente una visione integrale dei rating forniti dai clienti.

Gli istogrammi mostrati di seguito sono interattivi al fine di fornire una fruizione migliore da parte dell'utente.

Reviewer Nationality

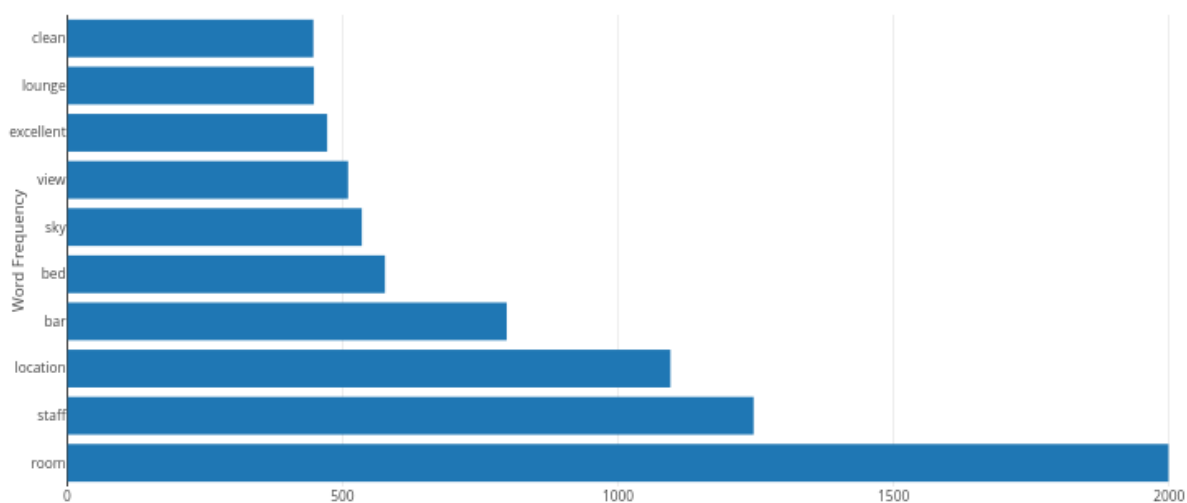


Review Average Score Distribution



inoltre viene visualizzato un grafico che mostra quali parole sono presenti più frequentemente nelle recensioni pre processate, tali parole sono utili nella ricerca con filtri mostrata di seguito al fine di filtrare il contenuto del dataset con le parole più frequenti:

Most Common Words



Interazione con il dataset

Dopo una prima analisi puramente descrittiva viene fornita una visione tabellare del dataset per consentire l'interazione dell'utente con lo stesso.

In questo modo si concede all'utente la possibilità di ordinare o filtrare i dati in base alle proprie esigenze:

	Reviewer_Score	Reviewer_Nation	review	sentiment
filter data...				
x	8.8	United States of America	Very nice lobby area Taxis right at door for transportation Helpful staff Good location for circle and district lines Beautiful top floor view of London	1
x	8.8	Ireland	Breakfast was a bit pricy and the excellent Complimentary condime	1
x	9.0	Ireland	Excellent location very chic hotel	1
x	7.1	United Kingdom	The fact that we were charged 12.5	0
x	9.0	Ireland	Nothing really The freshly bal	1
x	9.2	United Kingdom	It s v v v expensive which is hard to justify given there are cheaper hotels nearby Lovely hotel have stayed here twice and fab service and experience both times Great location Fabulous rooftop bar	1
x	9.0	Canada	The sky bar is the most expensive bar I ve ever seen Location is great A two minute walk from Tower Hill tube station Breakfast has every option you could possibly want	1
x	5.4	Canada	Our room was clean but the common areas hallways seemed run down the concierge was also not the most helpful Easy to access the tube and many local sights	1
x	9.0	United States of America	Concierge was extremely helpful and full of positive energy Tony and his colleague helped collect a box so I could FedEx items back to the U S all I had to do was pack and provide my account number and address THANK YOU They were	1
x	6.3	India	The worst experience was at the sky bar lounge where it felt that I am being paid to have a drink The attitude of the bar tender was not at all acceptable I had expressed checked out and have not received my invoice yet Its been 5	1

« < 1 / 322 > »

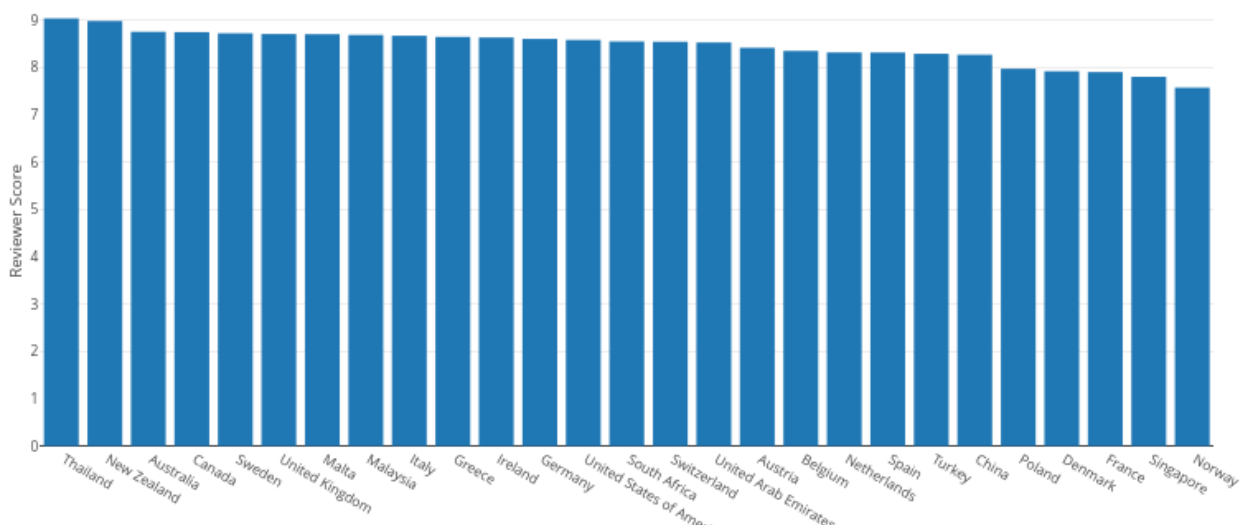
La componente mostrata consente le seguenti operazioni:

- Ordinare in ordine crescente o decrescente e ricercare nel dataset secondo diversi criteri combinabili:
 - Reviewer Score
 - Nazionalità
 - Recensione
 - Sentiment Ottenuto
- Escludere una recensione dalla virtualizzazione, utile nel caso in cui siano presenti recensioni per esempio diffamatorie che non hanno rilevanza
- La possibilità di ricerca per testo di una recensione

I filtri che possono essere applicati sopra permettono di cercare solo recensioni specifiche; per esempio, tutte le recensioni con la parole 'room' e di nazionalità 'italy', tale criterio verrà utilizzato per costruire grafici dinamici: in questo caso verranno visualizzati solo i grafici relativi alle recensioni che contengono la parola 'room' e sono di cittadini italiani.

In questo modo si ottengono grafici dinamici che evolvono rispetto ai filtri applicati.

Top Raters by Country



Il seguente plot mostra il review score medio ordinato in modo decrescente in base alla Nazionalità dei recensori.

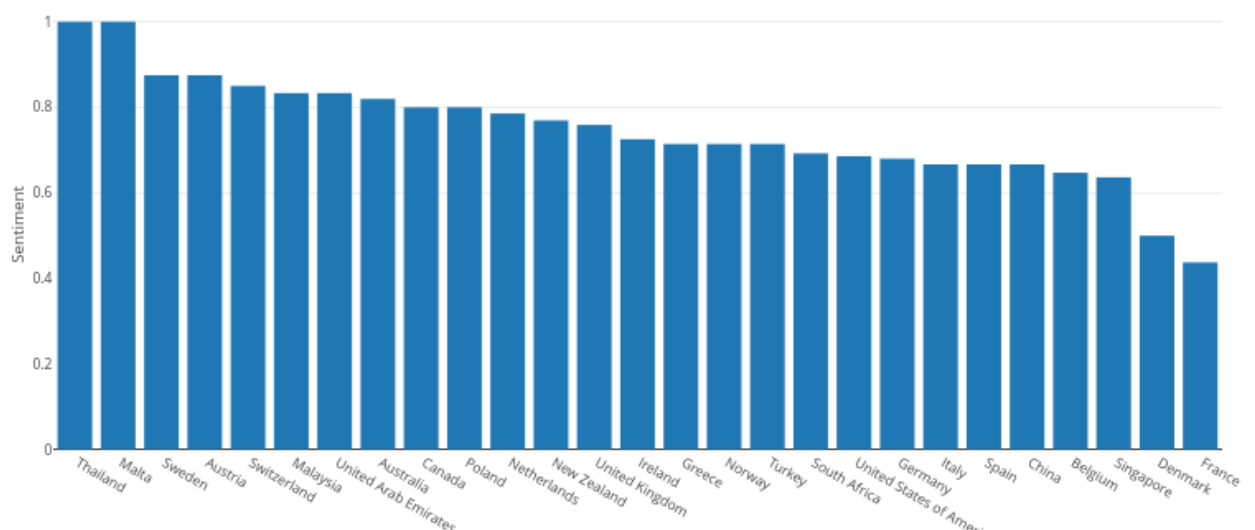
Nel grafico vengono considerati solo Stati che contano almeno 5 recensioni, in modo da ottenere evidenza statistica.

Non si ritiene valido un review score medio ottenuto da una sola recensione, 5 recensioni possono rappresentare - seppure parzialmente - l'opinione di uno stato.

La visione speculare del grafico illustrato viene fatta con la sentiment analysis: nell'istogramma viene mostrato in percentuale rispetto alla Nazionalità, quante recensioni, in media, sono state classificate positivamente dall'analisi dei sentimenti.

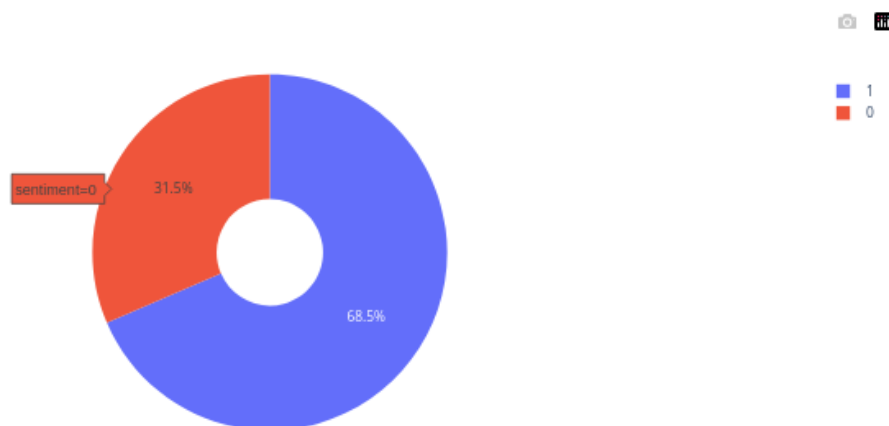
Per esempio un valore pari a 33% indica che il 33% delle recensioni di quella nazionalità è stato classificato, positivamente mentre il restante come negativo.

Percentage of Positive Sentiment by Country



Di seguito si riporta l'ultimo grafico dinamico che è stato implementato; questo mostra la distribuzione della sentiment - positiva o negativa - filtrando le recensioni secondo un determinato criterio. Di seguito viene riportato un esempio il quale riporta i risultati ottenuti filtrando le recensioni con la parola "room":

Sentiment Distribution



Osservando il grafico è possibile notare che, rispetto al numero totale di recensioni che contengono tale parola, il 68% di queste sono positive mentre le restanti sono negative. Ciò consente di avere un'idea di come la camera venga percepita dall'utente.

Explanation

Di seguito vengono mostrate le componenti necessarie per l'explination del dataset tramite SHAP e LIME.

Passando il cursore su una recensione è possibile leggerla nell'interezza.

Hotel Review Explanation

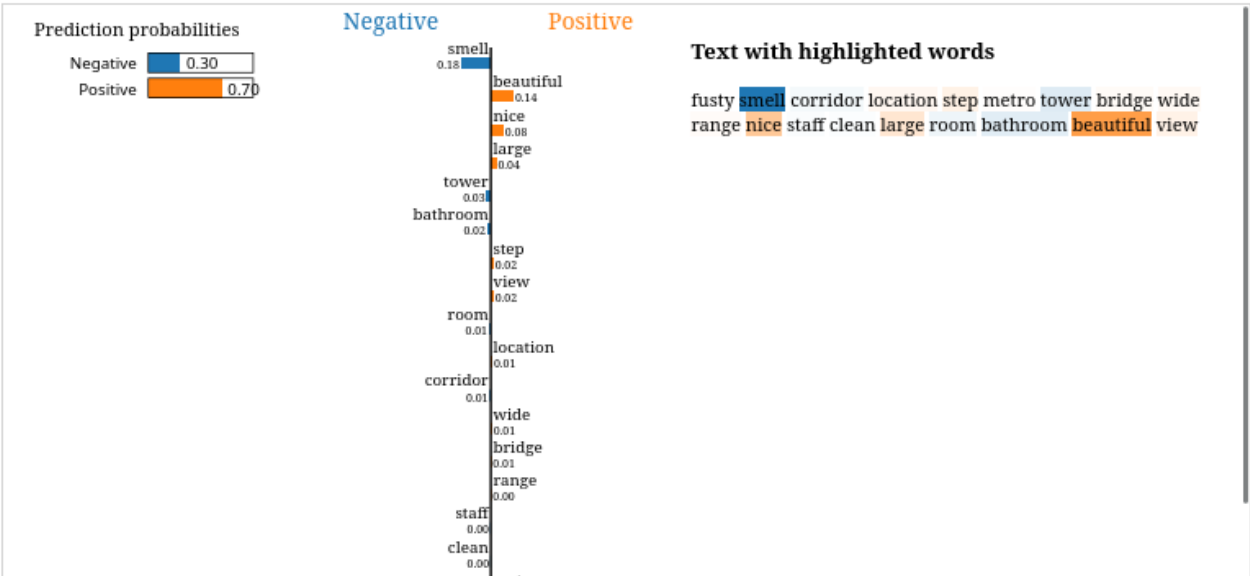
filter data...

Reviewer_Score	Reviewer_Nation	
5.4	Canada	Our room was clean but the common
9.0	United States of America	Concierge was extremely helpful FedEx items back to the U S all
8.8	United Kingdom	The picture of the rooms are de very small for 2 Our room looked
9.0	United States of America	Nothing Everything was quite g restaurant inside hotel wi
8.3	United Kingdom	No bath in the room Staff were ver
7.5	United Kingdom	the view from our room looked str but we were obviously just unluck
10	United Kingdom	Drinks quite expensive The complimentary water The breakf
5.8	Ireland	Arrived early room not ready to ready and it would be over an
9.2	United Kingdom	Poor coffee as usual and long wait to check in although this was down to extra security measures Well appointed rooms including Apple TV Modern bathrooms and large rooms
10	United Kingdom	No poached eggs available they had all gone We asked for a disabled access room but didn t get one I went to the front desk reception they put us in a disabled access room Absolutely perfect

Concierge was extremely helpful and full of positive energy Tony and his colleague helped collect a box so I could FedEx items back to the U S all I had to do was pack and provide my account number and address THANK YOU They were also extremely knowledgeable of the city and surrounding areas Provided excellent The concierge provided excellent transportation directions to various areas Again THANKS Front desk was aware of my arrival from the another country and had a room available for me to check in upon my arrival from Manchester The gal at the front desk was very hospitable beautiful smile and suggested a great place for a traditional English breakfast All the staff I met are keepers

review	sentiment
also not the most helpful Easy to ess the tube and many local sights	1
ue helped collect a box so I could er and address THANK YOU They were	1
than I expected the queen bed was ise from building work opposite We	0
friendly staff well designed nice s a natural and relaxed atmosphere	1
Room was compact but adequate for our stay Roof top bar was great	1
m sure some people got good views tc Warm cookie on arrival was also	1
lovely shower nice toiletries and lent variety of food and the staff	1
lem Went back after 15 00 room not e got to room half eaten chocolate	0
	1
	1

Cliccando sulla recensione, di lato verrà mostrata l'explication local tramite LIME. Per esempio:



Tale componente consente di leggere la recensione e capire quali parole hanno avuto un'influenza negativa o positiva.

In questo caso, l'odore nel corridoio influenza negativamente per la parola 'smell'; mentre, altre parole influenzano positivamente, come 'nice' e 'beautiful'.

Questo strumento consente di comprendere perché le recensioni vengono classificate positivamente o negativamente, e con quale confidenza: in questo caso, classificando correttamente, si considera una confidenza di recensione positiva al 70%.

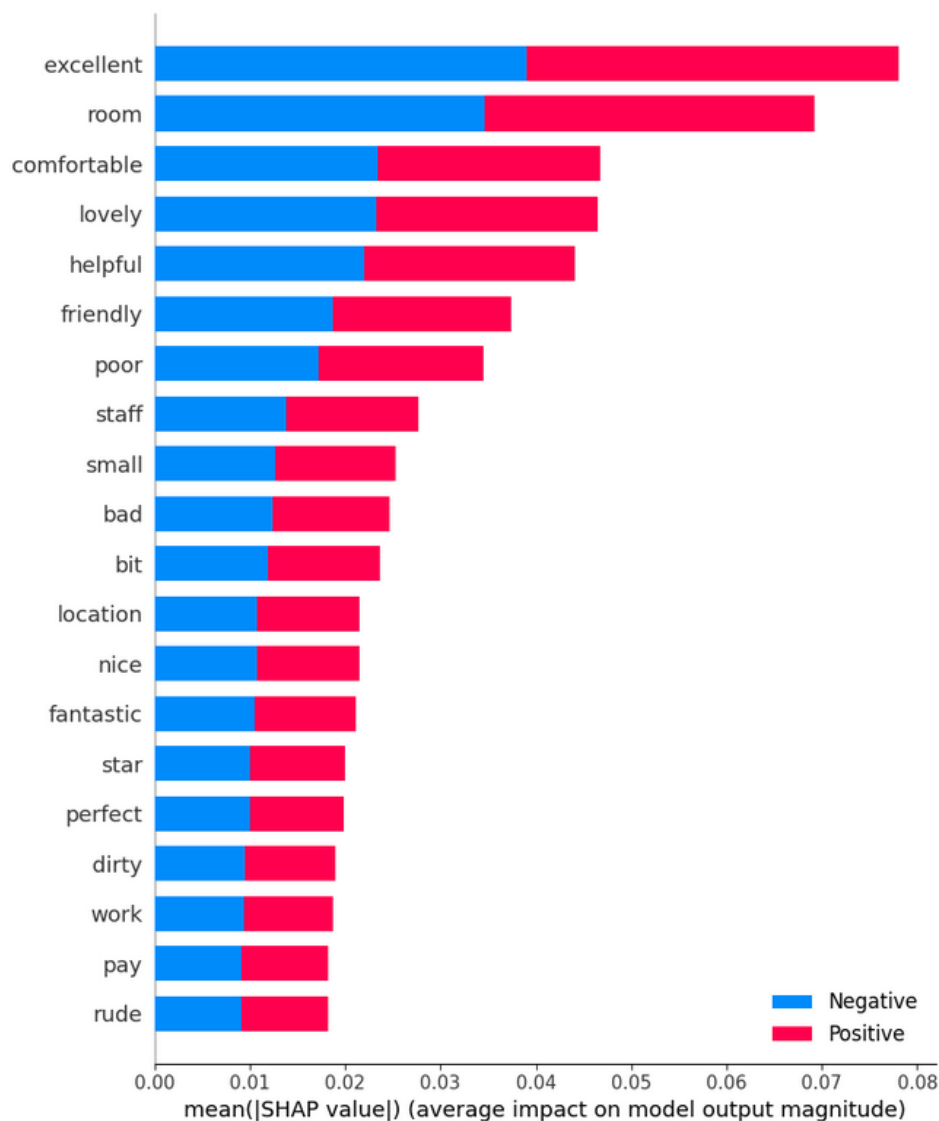
Ciò consente una explanation locale.

Il framework dash genera anche i grafici SHAP relativi al dataset delle recensioni del nostro hotel. Per motivi legati all'integrazione, questi non possono essere riportati direttamente nella Dash, ma vengono inseriti come immagini e visualizzati nella finestra del browser di fianco alla dash.

Tali grafici possono essere consultati per identificare i punti di forza e debolezza della struttura: parole che influenzano maggiormente o negativamente la predizione.

Inoltre, possono essere utilizzati in combinazione al filtraggio del dataset per cercare di analizzare con granularità più fine l'impatto di una parola sulla singola recensione

Per esempio, il grafico mostrato è il seguente; ma vengono prodotti anche i grafici mostrati precedentemente relativi al train set:



Infine per adempiere a quanto posto negli obiettivi iniziale come ‘Scenario Customer’, si propone un tool da fornire all’utente, per esempio integrato in una piattaforma di recensioni come Booking, per permettere all’utente di capire se sta esprimendo effettivamente quanto vuole rispetto alla struttura visitata. Ciò consente di migliorare la consapevolezza dell’utente nella scrittura di una recensione e quale impatto stanno avendo le parole che sta utilizzando.

Supponiamo il caso in cui l’utente si sia trovato bene nella struttura e abbia trovato camera pulita e staff cordiale, mentre non abbia apprezzato molto la colazione.

Quindi l’utente vuole esprimere una recensione positiva in generale ma mostrando che la colazione non è stata molto gradita, nonostante sia comunque stata accettabile.

Pertanto, una volta che l’utente avrà espressa la propria opinione, il sistema restituirà un feedback rispetto a quanto scritto per capire se sta descrivendo correttamente il proprio grado di soddisfazione.

Quindi durante la scrittura l’utente riceverà costanti input tramite LIME explanation; il tool proposto fornisce feedback solamente se la recensione ha almeno 10 caratteri.

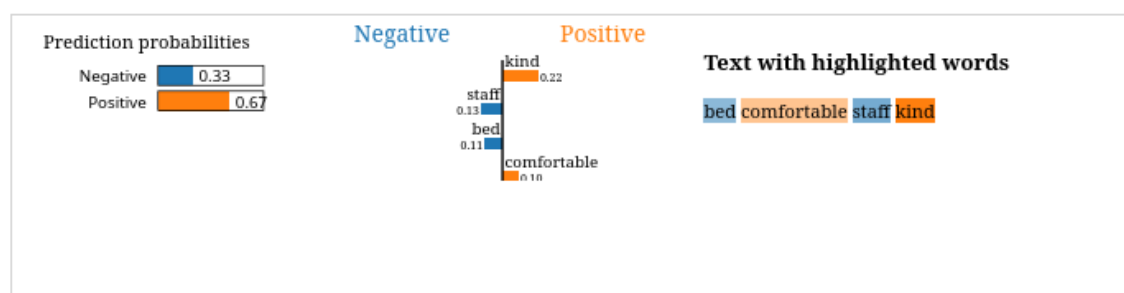
Si riporta di seguito un esempio di utilizzo, durante il quale si scrive la prima metà della recensione positiva e la seconda negativa.

Si riporta prima il risultato ottenuto con la prima metà:

Review Sentiment Explanation

The bed was comfortable and the staff very kind.

That's GREAT! 1



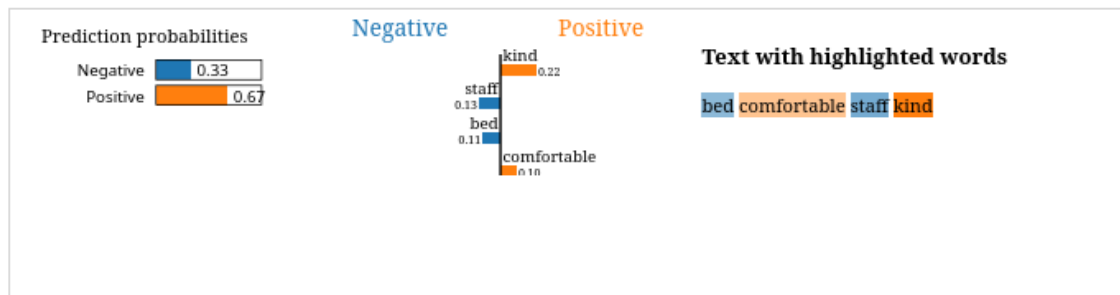
Viene indicato che in linea generale si sta esprimendo un giudizio positivo come aspettato.

Ora si riporta il risultato ottenuto aggiungendo la parte meno positiva cercando di capire come scriverla per non dare troppo impatto, in quanto la colazione è ritenuta accettabile.

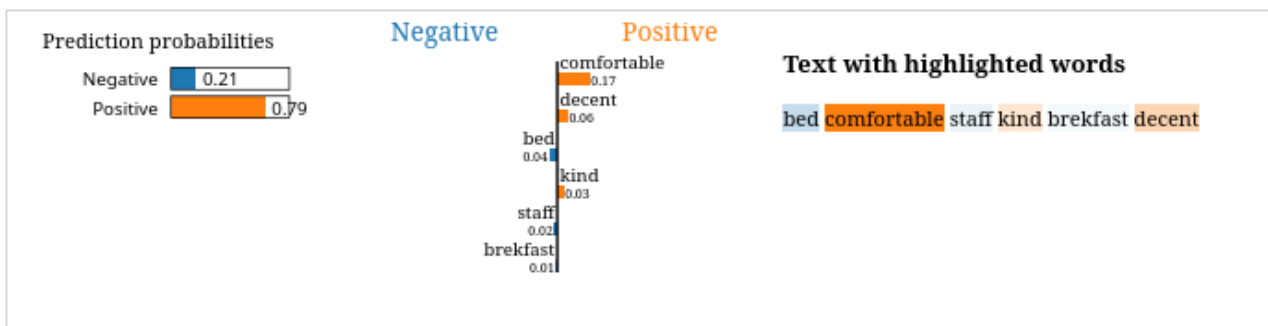
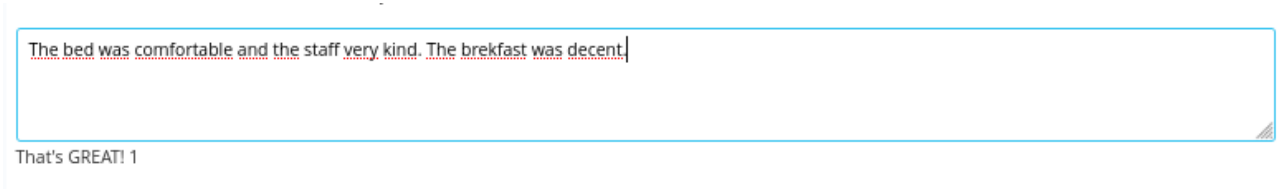
In questo caso aggiungendo alla recensione ‘The breakfast was pretty bad’, nonostante si stia esprimendo che la colazione non era ottimale, si ottiene un sentiment diverso, in quanto si utilizza una parola con accezione completamente negativa ‘bad’.

The bed was comfortable and the staff very kind. The breakfast was pretty bad.

Too much Negativity! 0



Quindi, per esprimere al meglio la propria opinione, risulta più adatto usare al posto di 'pretty bad' il termine 'decent'. Ottenendo il seguente risultato:



In questo modo è stato espresso un feedback positivo, comunque evidenziando alcune carenze sulla colazione.

L'utente in questo modo acquisisce consapevolezza sulle parole da utilizzare per esprimere chiaramente le proprie opinioni dando il giusto peso ed impatto alle parole utilizzate.

Il tool presentato potrebbe migliorare con un modello allenato a riconoscere un numero maggiore di parole e utilizzare feedback con granularità più fine.

Scenari

Hilton Hotel London

Si sceglie di utilizzare questa porzione di dataset, non utilizzata nel train e nel test del modello, poiché contiene un numero significativo di recensioni (circa 3000) al fine di testare la dashboard proposta.

Un numero significativo di recensioni consente di avere evidenza statistica su quanto analizzato. Alla dashboard viene passato il dataset contenente solo le recensioni di questo specifico hotel.

Si ottiene che, in generale, la sentiment analysis è concorde con i review score indicati dal modello, con un errore del 20% cento.

Infatti otteniamo:

Sentiment Positive Review: 75.34%

Dataset Positive Review: 93.59%

Circa il 20% delle recensioni viene classificata con sentiment positivo nonostante sia negativo, ciò può essere causato dai limiti del modello di training allenato con poche feature (viste le risorse disponibili).

Il resto della dashboard consente di estrarre gli insight sopra descritti, cercando di analizzare il grado di gradimento dei vari servizi offerti dall'hotel.

La dashboard produce in output l'importanza delle feature SHAP mostrando quelle più rilevanti. Esse potrebbero corrispondere ad aspetti rilevanti della struttura, le quali potrebbero impattare positivamente o negativamente sulla stessa.

Inoltre, analisi più granulari sui dataset consentono di estrarre informazioni rilevanti sui servizi da migliorare o i servizi apprezzati dagli utenti.

Dataset ChatGPT

Quanto proposto è un approccio sperimentale - visti i limiti attuali di ChatGPT. Tale test ci consente di valutare il modello su recensioni provenienti da un contesto completamente diverso da quello di test e di train.

Nonostante la varietà del lessico utilizzato e la diversa struttura delle recensioni, i risultati ottenuti sono buoni: nella maggior parte dei casi, il review score previsto dal modello, coincide con lo score espresso, visibile nella parte iniziale della Dash.

Le recensioni generate da ChatGPT sono molto semplici ma significative, per esempio:

- 8, "I had a great stay at this hotel. The room was spacious and well-maintained. The staff was friendly and attentive. Highly recommended."
- 5, "I was disappointed with my stay at this hotel. The room was dirty and had a strange smell. The staff was not helpful either. Would not stay here again.", India
- 9, "Excellent hotel with a beautiful view. The room was clean and comfortable. The staff went above and beyond to make our stay enjoyable.", United States

Dalla dash si ottiene, considerate solo le 41 recensioni generate nel dataset, che la sentiment è d'accordo con il rating del recensore. Infatti si ottiene:

- Sentiment Positive Review: 73.17%
- Dataset Positive Review: 73.17%

Per il resto del dataset valgono le considerazioni precedenti.

Quindi, il test effettuato sul dataset generato artificialmente tramite Chat GPT consente di effettuare ulteriori analisi sul modello proposto. I risultati ottenuti su tale dataset sono buoni, nonostante il trainset sia completamente diverso e non specifico per hotel di lusso.

Considerando questa prima analisi, il modello sembra funzionare bene anche su dataset diversi. Nonostante ciò, sarebbe opportuno effettuare ulteriori esperimenti con altri set di dati a conferma di quanto detto.

In conclusione, la dashboard proposta consente di fornire insights alla struttura mostrando aspetti che impattano maggiormente e informazioni relative alla clientela suddivise per nazionalità.

Analisi Finali

In seguito alle analisi precedentemente svolte alla presentazione della dashboard, si possono trarre conclusioni sul lavoro svolto.

Insights

Il lavoro svolto fornisce uno strumento in grado di estrarre insights significativi utili per una struttura ricettiva al fine di prendere recensioni data driven.

Obiettivi Iniziali

Le domande poste durante la presentazione degli obiettivi, ottengono una risposta tramite il lavoro svolto. In particolare:

Nello **scenario business**, relativo alla struttura alberghiera:

‘Quali aspetti devono essere migliorati ? Quali sono i punti di forza della struttura?’

L'utilizzo di SHAP consente di ottenere un feedback sulle parole, quantificando la loro influenza nella predizione di ciascuna recensione.

Ciò nonostante, non si ottengono parole che influiscono solo negativamente o solo positivamente, il sentiment complessivo dipende dalla combinazione dell'influenza di più parole.

Lo SHAP value associato ad ogni parola può essere utilizzato - nelle analisi svolte sulla dash - per fornire indicazione su aspetti da migliorare o no della struttura.

‘La clientela di diversa nazionalità ha diverse preferenze?’

La dash proposta consente di analizzare le preferenze della clientela di diverse nazionalità al fine di massimizzare la soddisfazione del cliente.

Lato customer, si possono raccogliere le seguenti osservazioni:

‘Il cliente sta esprimendo correttamente quanto vuole intendere riguardo all'hotel?’

Il tool presente nella dash fornisce un Feedback real-time durante la scrittura individuando le parole più efficaci: così facendo assegna il giusto peso a tutti i vocaboli e fa in modo che la recensione rispecchi il più possibile il proprio giudizio.

Riguardo l'obiettivo comune:

‘Lo score è discordante con quanto espresso nella recensione testuale?’

Il riscontro è che potrebbe esserlo, ma per valutare tale aspetto è necessaria un'analisi manuale, supportata dalla dashboard proposta.

Recensioni con score discordanti possono impattare sul rating generale della struttura, pertanto, individuarle è di fondamentale importanza.

La sentiment analysis consente di analizzare ogni aspetto della recensione e individuare quelle incoerenti con il Review Score assegnato.

Tali recensioni manualmente analizzate possono poi essere rimosse dal dataset, o, eventualmente, si può richiedere una revisione all'utente.

Un esempio è la recensione "everything excellent" con score assegnato pari a 2.

Campioni simili potrebbero alterare i modelli di apprendimento e/o avere impatti negativi sulla struttura.

Lunghezza Recensioni

Tra gli insights ottenuti, vi è quello relativo alla lunghezza delle recensioni. Come atteso, le recensioni negative sono più lunghe delle positive: il pubblico tendenzialmente tende ad essere proliquo nella descrizione di aspetti negativi rispetto alla descrizione di aspetti positivi.

Il che potrebbe risultare utile al fine di indagare maggiormente sugli aspetti che non sono stati apprezzati dalla clientela, in modo da poter intervenire per migliorare il servizio.

Limiti

Il lavoro svolto, nonostante i risultati ottenuti, presenta alcuni aspetti limitanti:

- Sono stati utilizzati **vettorizzazioni context-free** che non tengono in considerazione l'ordine e la posizione delle parole nella frase analizzando le parole come indipendenti, ciò limita notevolmente l'analisi dei testi non individuando alcuni aspetti come sarebbe possibile utilizzando modelli come word2vec. L'utilizzo di modelli di vettorizzazione complessi potrebbe migliorare le performance dei modelli tenendo conto del contesto, a discapito dell'explainability, il quale risulta un punto chiave del lavoro svolto.
- Le analisi svolte e i modelli prodotti sono stati realizzati in base alle **risorse computazionali limitate** rese disponibili da GoogleColab.
Infatti, è stato necessario svolgere un ridimensionamento del dataset, andando a ridurre notevolmente la quantità di dati contenuti nella classe positiva.
Inoltre, non viene eseguito alcun tuning manuale dei modelli: un migliore tuning consentirebbe l'ottenimento di modelli più performanti con parametri ottimali.
Infine, la vettorizzazione scelta, opera solo su parole con un'occorrenza maggiore di 100 nel dataset: parole poco usate non vengono considerate, pertanto il modello potrebbe non riconoscere alcune parole presenti nel corpus.

Nel progetto ci si è concentrati sul produrre insight significativi rispetto alla performance dei modelli.
- Il modello è allenato solo su recensioni in inglese, scritte da clienti di diverse nazionalità che hanno alloggiato in **hotel di lusso in Europa**.
Il lessico che verrà riconosciuto dai modelli sarà limitato a una clientela abbiente che possa permettersi di alloggiare in una struttura dal costo elevato. Pertanto, un lessico più semplice e meno dettagliato utilizzato da clientela giovane potrebbe non essere riconosciuto. Sono necessarie ulteriori analisi per valutare la scalabilità del modello.
- Il modello è allenato solo tramite recensioni ottenute su **una piattaforma Booking.com**, quindi potrebbe essere limitato per esprimere il sentiment relativo ad una struttura utilizzando solo feedback provenienti da una piattaforma specifica. Tecniche di Data Integration permetterebbero una rappresentazione più significativa

Conclusioni

L'obiettivo del progetto viene raggiunto generando dati utili in ambito decisionale, vengono forniti insight significativi utili per intraprendere scelte all'interno di strutture alberghiere relative ad aspetti da migliorare e servizi da potenziare.

Infatti, la dashboard proposta e i risultati ottenuti consentono di analizzare al meglio le recensioni estraendo informazioni utili alle decisioni.

Per esempio, ottenendo feedback spesso negativi relativi alle camere, è possibile analizzare le varie recensioni per ottenere informazioni utili al fine di comprendere cosa migliorare di questo aspetto.

Pertanto, i dati vengono elaborati ed utilizzati per generare grafici ed effettuare analisi utili a capire il grado di soddisfazione dell'utente rispetto alla struttura alberghiera e ai servizi offerti.

Inoltre, il lavoro svolto ha permesso di fornire una maggiore consapevolezza rispetto all'utilizzo delle parole nelle recensioni e al loro impatto nell'esprimere quanto voluto dell'utente.

Le performance ottenute dal modello, considerati i limiti esposti precedentemente, sono buone sia nelle metriche presentate che nell'utilizzo dei modelli nella dashboard.

Un possibile sviluppo futuro è quello di testare i modelli utilizzando una tecnica di vettorizzazione differente che consideri un numero maggiore di parole, e costruisca delle pipeline in grado di effettuare continuo retraining in base alle nuove recensioni ottenute.

In conclusione, il lavoro svolto ha permesso di raggiungere gli obiettivi proposti estraendo insights significativi relativi a strutture alberghiere per dare valore alle recensioni lasciate dai clienti.

Considerato l'impatto delle recensioni è di fondamentale importanza tenere in considerazione l'opinione del cliente al fine di ottenere la massima soddisfazione e di conseguenza massimizzare i profitti.