

*Ministero dell'istruzione e del merito***H038 - STAATLICHE ABSCHLUSSPRÜFUNG DER OBERSCHULE****Fachrichtung** ITIA – INFORMATIK UND TELEKOMMUNIKATION**SCHWERPUNKT** INFORMATIK**Fach:** INFORMATIK***Führen Sie Teil I der Arbeit und zwei der vorgeschlagenen Fragestellungen aus Teil II aus.*****TEIL I****Prämisse**

Ein *dataset* (wörtlich: „Menge von Daten“) ist eine strukturierte Sammlung einer großen Menge von Daten; im untenstehenden Fallbeispiel ist das *dataset* in relationaler Form organisiert. In dieser Datensammlung werden Elemente der realen Welt beschrieben (z. B. Ereignisse, Nachrichten, Objekte usw.), jedes mit einer Liste von Merkmalen.

Ein *dataset* kann Zahlen, Wörter, Bilder, Töne oder jegliche andere Art von Informationen beinhalten.

Datasets sind entscheidend für das Training bestimmter Anwendungen der künstlichen Intelligenz. Der Vorgang des *labeling* (Beschriftung) eines *datasets* bedeutet das Anbringen von *labels* (Beschriftungen) an jedem Element, um aufzuzeigen, was dieses darstellt, oder besser, zu welcher Kategorie oder Klasse es gehört. Stellen Sie sich zum Beispiel ein *dataset* mit Bildern von Blumen vor. Das *labeling* ordnet jedem Bild eine Beschriftung zu, also eine Zeichenkette mit dem Namen der Blumenart, welche auf dem Bild abgebildet ist. Die Aufgabe des *labelings* wird normalerweise von Personen durchgeführt, welche die Klassifizierung der Elemente des *datasets* durchführen können, in diesem Beispiel ein Botaniker. Diese Operation ist notwendig, wenn man beabsichtigt, Techniken der künstlichen Intelligenz und des *machine learning* basierend auf Algorithmen für überwachtes Lernen zu verwenden.

In diesem Fall wird ein *dataset* mit einer großen Anzahl an bereits beschrifteten Elementen vorbereitet, das sogenannte *training-dataset*, welches eine Menge an Beispielen umfasst, um den Algorithmus zu trainieren (oder richtiger gesagt, um das Modell der künstlichen Intelligenz zu trainieren). Wenn der Algorithmus die zur Verfügung gestellten Beispiele „gelernt“ hat, wird er in der Lage sein, autonom neue Elemente zu klassifizieren. Wenn zum Beispiel als Input ein neues Bild einer Blume gegeben wird, sollte der trainierte Algorithmus in der Lage sein, mit großer Wahrscheinlichkeit die Klassifizierung zu liefern, zum Beispiel „Margerite“.

Fallbeispiel

Um dem Phänomen der *fake news* entgegenzuwirken, wurde einem Informatik-Unternehmen der Auftrag erteilt, eine Web-Plattform zu entwickeln, um das *labeling* eines *training-datasets* mit großen Dimensionen durchzuführen. Mit diesem *dataset* soll ein Modell der künstlichen Intelligenz trainiert werden, um *news* im Web zu klassifizieren.

Jede *news* wird durch die Quelle gekennzeichnet, aus der sie stammt. Von der Quelle wird die Art (Blog, soziale Medien, Online-Zeitung, Streaming-Plattform usw.) sowie der Name (New York Times, Gazzetta del Mezzogiorno, Focus, Facebook, Instagram, TikTok, Spotify, YouTube usw.) oder in jedem Fall die Domain der Herkunftsseite angegeben. Jede *news* ist außerdem durch eine URL gekennzeichnet, über die sie im Internet aufgefunden wird, durch ein Veröffentlichungsdatum, einen möglichen Titel, den Autor (falls verfügbar), den Textinhalt (Text von einem Zeitungsartikel, dem Beitrag in einem sozialen Netzwerk, einem Transkript eines Videos oder Podcasts usw.). Mit ihnen können möglicherweise mehrere Multimedia-Inhalte (Audio, Video und Bilder) sowie mehrere Kommentare verbunden sein, die zu den *news* gehören.

*Ministero dell'istruzione e del merito***H038 - STAATLICHE ABSCHLUSSPRÜFUNG DER OBERSCHULE****Fachrichtung** ITIA – INFORMATIK UND TELEKOMMUNIKATION

SCHWERPUNKT INFORMATIK

Fach: INFORMATIK

Das Ziel besteht darin, jede *news* zu klassifizieren und ihr zwei Labels zuzuweisen: *topic* und *result*. Die Beschriftung *topic* dient dazu anzugeben, auf welches Thema sich die *news* bezieht, und kann einen Wert aus einer Liste annehmen wie beispielsweise Wirtschaft, Politik, Medizin und Gesundheit, Kultur, Nachrichten, Wissenschaft und Technologie, Sport usw.

Die Beschriftung *result* ist diejenige, welche die *news* tatsächlich klassifiziert und einen der folgenden möglichen Werte zuweist: „Fake“ oder „Wahr“ oder „Zweifelhaft“. Für den Fall, dass die *news* als *fake* eingestuft wird, muss außerdem einer der folgenden möglichen Gründe festgestellt werden: „erfundener Inhalt“, „manipulierter Inhalt“, „von Betrügern verbreiteter Inhalt“, „falscher Kontext“, „irreführender Inhalt“, „falscher Zusammenhang“, „Satire oder Parodie“ sowie anschließend eine Anmerkung zur Begründung eingefügt werden.

Die Tätigkeit des *labeling* wird einer Gruppe von Junior-Experten und einer Gruppe von Senior-Experten übertragen, die in zwei aufeinanderfolgenden Phasen tätig werden: In der ersten Phase führt die Gruppe von Junior-Experten eine vorübergehende Kennzeichnung der *news* durch; anschließend führt die Gruppe der Senior-Experten die endgültige Prüfung der in der ersten Phase durchgeführten Klassifizierung durch, wobei die Beschriftungen unverändert bleiben oder mit den von den Junior-Experten gegebenen entsprechenden Begründungen korrigiert werden.

Während des *labelings* gibt es somit *news*, die noch nicht gekennzeichnet wurden, *news*, die eine vorläufige Beschriftung erhalten haben, und *news* mit geprüfter Beschriftung.

Am Ende der Beschriftung des *datasets* sollte die Plattform auch eine Analyse der beschrifteten Daten ermöglichen.

Erfüllen Sie folgende Arbeitsaufträge, nachdem Sie die entsprechenden zusätzlichen Annahmen getroffen haben:

1. Erstellen Sie eine Analyse der Referenzrealität, die zur Definition des konzeptionellen Schemas der relationalen Datenbank führt, das Ihrer begründeten Meinung nach für die Verwaltung der Webplattform geeignet ist
2. Erstellen Sie das logische Schema der relationalen Datenbank
3. Definieren Sie in der Sprache SQL eine Teilmenge der Relationen der Datenbank, die einige der Relationen enthält, die referentielle Integritätsbeschränkungen und/oder Domänenbeschränkungen enthalten, sofern vorhanden
4. Entwickeln Sie in der Sprache SQL Abfragen, um die folgenden Informationen zu erhalten:
 - a) Die Anzahl der als gefälscht klassifizierten *news* (*fake*), für jeden Wert der Beschriftung *topic* welche im vergangenen Jahr veröffentlicht wurden
 - b) Die Anzahl der *news*, die von einem bestimmten hochrangigen Experten analysiert wurden, dessen Vor- und Nachname angegeben wurde

*Ministero dell'istruzione e del merito***H038 - STAATLICHE ABSCHLUSSPRÜFUNG DER OBERSCHULE****Fachrichtung** ITIA – INFORMATIK UND TELEKOMMUNIKATION**SCHWERPUNKT** INFORMATIK**Fach:** INFORMATIK

- c) Für ein gegebenes Thema (*topic*) und einen Veröffentlichungszeitraum die Anzahl der als *fake* eingestuften *news* für jeden Grund („erfundener Inhalt“, „manipulierter Inhalt“, ..., „Satire oder Parodie“)
 - d) Den Namen des Senior-Experten, der im letzten Monat die meisten Prüfungen der „gefälschten“ *news* durchgeführt hat
5. Entwerfen Sie die Grobstruktur einer Webanwendung für die Verwaltung der oben dargestellten Realität, wobei anzugeben ist, welche technologischen Ressourcen, sowohl Hardware als auch Software, für ihre Verwirklichung geeignet sind
6. Realisieren Sie einen Teil der Anwendung, sowohl Client- als auch Serverseite, unter Verwendung geeigneter Sprachen Ihrer Wahl, welche eine der folgenden Funktionalitäten erfüllt:
- a) Darstellung aller Daten einer bestimmten, noch nicht beschrifteten *news* für den Junior-Experten, um eine vorläufige Beschriftung zu ermöglichen
 - b) Darstellung, möglicherweise in grafischer Form, der Anzahl der als *fake* gekennzeichneten *news* für jedes Thema (*topic*);
 - c) Formular, das es dem Benutzer ermöglicht, ein Thema und einen Veröffentlichungszeitraum auszuwählen und die Anzahl der als *fake* eingestuften Nachrichten für jeden Grund anzuzeigen.

TEIL II

- I. In Bezug auf das im Teil I vorgestellte Thema wird angenommen, dass das Unternehmen jede der beiden Beschriftungsphasen überwachen muss. Erörtern Sie die möglichen Änderungen an den Funktionalitäten der Plattform, um die Anzahl der täglich von jedem Experten beschrifteten *news* und die Gesamtzeit, die jede Gruppe von Experten (Junior und Senior) für die Beschriftung des gesamten *datasets* benötigt, zu verfolgen.
- II. Nehmen Sie in Bezug auf das im Teil I vorgestellte Thema an, dass mögliche Änderungen von Senior-Experten an den ursprünglich von Junior-Experten eingegebenen Beschriftungen verfolgt werden sollen. Beschreiben Sie die notwendigen Änderungen an der Datenbank für diese Anforderung.
- III. Nehmen Sie an, dass für den Zugang zu einer Website zusätzlich zu den klassischen Anmeldedaten (Benutzername und Passwort) eine weitere Ebene der Benutzerverifizierung (Zwei-Faktor-Authentifizierung) hinzugefügt werden soll. Beschreiben Sie eine mögliche Benutzertabellenstruktur zur Unterstützung dieses Authentifizierungsmodus. Beschreiben Sie auch die grundlegenden Schritte, des Quellcodes auf Server-Seite, um den Zugriff auf die Website zu gewähren.



Ministero dell'istruzione e del merito

H038 - STAATLICHE ABSCHLUSSPRÜFUNG DER OBERSCHULE

Fachrichtung ITIA – INFORMATIK UND TELEKOMMUNIKATION

SCHWERPUNKT INFORMATIK

Fach: INFORMATIK

- IV. Eine relationale Datenbank ermöglicht den gemeinsamen und gleichzeitigen Zugriff mehrerer Benutzer über das Netz. Beschreiben Sie die möglichen Probleme, die die Integrität und Konsistenz der Daten beeinträchtigen können und durch welche Maßnahmen und Lösungen sie verhindert werden können.

Dauer der Arbeit: 6 Stunden

Der Gebrauch von Referenzhandbüchern für Programmiersprachen (language reference) ist erlaubt.

Der Gebrauch wissenschaftlicher und/oder grafischer Taschenrechner ist erlaubt, sofern diese nicht mit einem CAS (Computer Algebra System) oder SAS (Symbolic Algebra System) ausgestattet sind und keine Verbindungsmöglichkeit zum Internet aufweisen.

Der Gebrauch eines deutschsprachigen Wörterbuchs ist erlaubt.

Der Gebrauch eines zweisprachigen Wörterbuchs (Deutsch - Sprache des Herkunftslandes) ist für Kandidatinnen und Kandidaten mit nicht deutscher Muttersprache erlaubt.

Das Schulgebäude darf erst drei Stunden nach Bekanntgabe des Themas verlassen werden.