

## **KIWaste - Modification Analysis of AI Algorithms in the waste management sector**

### **Project description:**

Nowadays, more and more, value is placed on sustainability. This also includes waste disposal. "Zentrale Abfallwirtschaft Kaiserslautern" (ZAK) is a private company that operates in the waste management sector. The company allows private persons to use their services once a day for free, commercial users on the other hand always have to pay. However, there is a significant number of companies who disguise themselves as private persons and dispose their business waste illegally for free or use the disposal service more often than allowed. ZAK is also a client of the management consulting company Trusted Decisions (TD), which analyses and visualises data with the business intelligence tool QlikSense.

The aim of this work is to examine existing AI algorithms implemented by default and determine to what extent they can be modified. Those algorithms were implemented in Python at the client's server and should ensure the most efficient identification of customers disposing illegally.

### **Procedure:**

For each disposal in the waste company, an entry is created in the company's database. These entries consist of 12 data fields, which contain information such as name, registration number, e-mail, and type of waste. The project team focused on two datasets. The first dataset was example data of the client's system, and the second data was validation data. The existing example data had to be analysed by the team and the validation data needed to be created based on information by the client. The used AI algorithms to identify the illegal customers are K-Nearest Neighbor, Isolation Forest, Last Outlier Factor and Deep Learning. To apply these algorithms with the customer data, knowledge had to be built up through a literature research. After that a parameter analysis was carried out and the in Python used parameters were identified, and possible changes defined.

Furthermore, a concept for the procedure and a protocol were created to be able to carry out uniform test runs with variations (regarding the value) of the parameters. This turned out to be particularly important because there are only two access points to the customer's server. To validate the data, the team created a QlikSense App with a data cleaning script and script to merge the result data of the algorithms with the validation data. This enabled the team to calculate measures in the QlikSense visualisation layer, which provide the accuracy of the given algorithms in comparison to the validation data.

### **Result:**

As a result of this analysis the project team created a concept to understand the given algorithms and their parameters. Additionally, the team analysed how the parameters of these algorithms can be changed, how those parameters influence the outcome of the algorithms, how the given data sets can be understood and used, and a set of validation data was created. All of this provides a concept to understand the algorithms and create a profound basis to build upon for further work with the given algorithms.

## **KI Waste – Veränderungsanalyse von KI-Algorithmen im Abfallwirtschaftssektor**

### **Projektbeschreibung:**

Heutzutage wird immer mehr Wert auf Nachhaltigkeit gelegt. Dazu gehört auch die Abfallentsorgungsfirma „Zentrale Abfallwirtschaft Kaiserslautern (ZAK)“. Das Unternehmen ermöglicht es Privatpersonen, ihre Dienste einmal täglich kostenlos zu nutzen, gewerbliche Nutzer müssen dagegen immer bezahlen. Allerdings gibt es eine nicht unerhebliche Zahl von Unternehmen, die sich als Privatpersonen tarnen und ihre Gewerbeabfälle kostenlos illegal entsorgen oder den Entsorgungsservice häufiger als erlaubt in Anspruch nehmen. ZAK ist zudem Kunde der Unternehmensberatung Trusted Decisions (TD), die Daten mit dem Business-Intelligence-Tool QlikSense analysiert und visualisiert. Ziel dieser Arbeit ist es, bestehende standardmäßig implementierte KI-Algorithmen zu untersuchen und festzustellen, inwieweit sie modifiziert werden können. Diese Algorithmen wurden in Python auf dem Server des Kunden implementiert. Diese Algorithmen sollen eine möglichst effiziente Identifizierung illegal entsorgender Kunden gewährleisten.

### **Vorgehensweise:**

Für jede Entsorgung im Entsorgungsunternehmen wird ein Eintrag in der Unternehmensdatenbank erstellt. Diese Einträge bestehen aus 12 Datenfeldern, die Informationen wie Name, Registrierungsnummer, E-Mail und Abfallart enthalten. Das Projektteam konzentrierte sich auf zwei Datensätze. Der erste Datensatz waren Beispieldaten des Kundensystems und der zweite Datensatz waren Validierungsdaten. Die vorhandenen Beispieldaten mussten vom Team analysiert und die erforderlichen Validierungsdaten basierend auf Informationen des Kunden erstellt werden. Die verwendeten KI-Algorithmen zur Identifizierung der illegalen Kunden sind K-Nearest Neighbor, Isolation Forest, Last Outlier Factor und Deep Learning. Um diese Algorithmen auf die Kundendaten anzuwenden, musste Wissen durch eine Literaturrecherche aufgebaut werden. Danach wurde eine Parameteranalyse durchgeführt und die in Python vorhandenen Parameter identifiziert und mögliche Änderungen definiert.

Weiters wurde ein Konzept und Protokoll erstellt, um einheitliche Testläufe mit Variationen (hinsichtlich des Wertes) der Parameter durchführen zu können. Dies erwies sich als besonders wichtig, da es nur zwei Zugänge zum Server des Kunden gibt. Um die Daten zu validieren, erstellte das Team eine QlikSense-App mit einem Skript zur Datenbereinigung und einem Skript, um die Ergebnisdaten der Algorithmen mit den Validierungsdaten zusammenzuführen. Dadurch konnte das Team Kennzahlen in der QlikSense-Visualisierungsschicht berechnen, die die Genauigkeit der gegebenen Algorithmen im Vergleich zu den Validierungsdaten liefern.

### **Ergebnis:**

Als Ergebnis dieser Analyse erstellte das Projektteam ein Konzept, um die gegebenen Algorithmen und ihre Parameter zu verstehen. Darüber hinaus analysierte das Team, wie die Parameter dieser Algorithmen geändert werden können, wie diese Parameter das Ergebnis der Algorithmen beeinflussen, wie die gegebenen Datensätze verstanden und verwendet werden können, und ein Satz von Validierungsdaten wurde erstellt. All das bietet ein Konzept und eine Grundlage, um auf den gegebenen Algorithmen aufzubauen um in zukünftigen Projekten eine hohe Genauigkeit zu erreichen.