



lezzgo

Missbrauchsdetektion in lezzgo

29.10.2018

Carlo Beltrame & Adrian Stucki



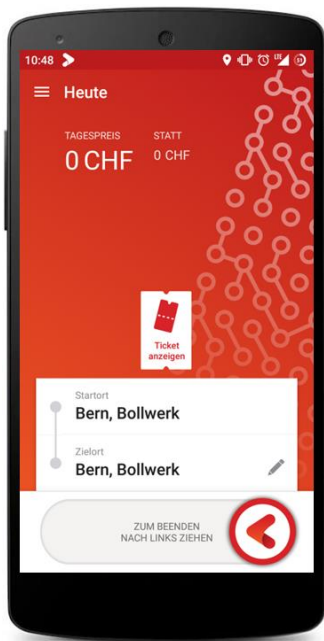


Agenda

- Ausgangslage
- Lösungsansatz
- Implementation
- Lessons learned
- Kurzer Ausblick



Was ist lezzgo?



Einsteigen

Mit einem Klick hast du dein gültiges Ticket in der Tasche

Reisen

Du musst dich nicht mehr mit verschiedenen Tarifen auskennen, kannst beliebig reisen und spontan umsteigen.

Aussteigen

Am Ende des Tages bezahlst du den optimalen Preis für die Strecke, die du gefahren bist.



Markttest automatisches Ticketing

Um mit lezzgo schweizweit
Tickets zu verkaufen...



...müssen wir Vorgaben
einhalten und rapportieren.



Missbrauch & Kontrolle



Preisberechnung



Reiseerfassung



Missbrauchsfälle öV im Vergleich

Missbrauchsszenarien	Ticketart	Risikobeurteilung	Beispiel
Out of Service-Ausrede	Papier		Automat defekt
	Prepaid-E-Ticket		Batterie leer
	lezzgo		Keine GPS Verbindung
Ticket nicht für gesamte Strecke gelöst	Papier		Teilstrecke lösen
	Prepaid-E-Ticket		Abschnittsweise lösen
	lezzgo		Später Check-In Früher Check-out
Manipulation des Trägermediums	Papier		Korrektur Gültigkeitsdatum
	Prepaid-E-Ticket		Screenshot
	lezzgo		Screenshot



Idee der BLS

- Mehrere 100 bis 1000 Reisen täglich



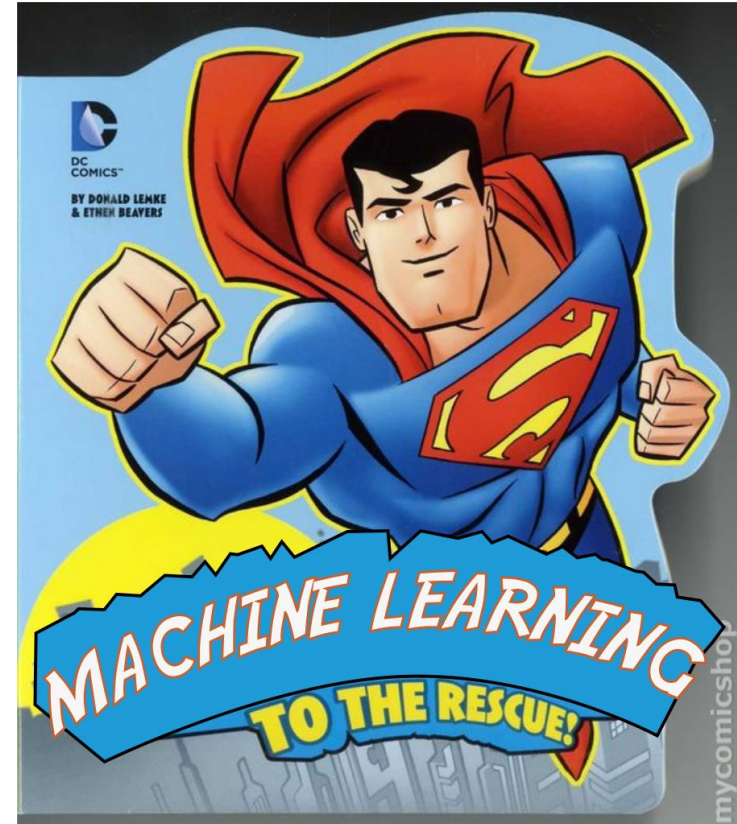
Idee der BLS

- Mehrere 100 bis 1000 Reisen täglich
- Manuelle Bewertung jeder Reise unmöglich



Idee der BLS

- Mehrere 100 bis 1000 Reisen täglich
- Manuelle Bewertung jeder Reise unmöglich





Lösungsidee

- Relevante Kennzahlen zu jeder Reise sammeln



Lösungsidee

- Relevante Kennzahlen zu jeder Reise sammeln
- Wie viel Zeit verging zwischen letzter Ticketkontrolle und Check-Out?



Lösungsidee

- Relevante Kennzahlen zu jeder Reise sammeln
- Wie viel Zeit verging zwischen letzter Ticketkontrolle und Check-Out?
- Wie schnell bewegte sich der Kunde während dem Check-Out fort?



Lösungsidee

- Relevante Kennzahlen zu jeder Reise sammeln
- Wie viel Zeit verging zwischen letzter Ticketkontrolle und Check-Out?
- Wie schnell bewegte sich der Kunde während dem Check-Out fort?
- Passen mehrere Reisen eines Kunden am gleichen Tag zusammen? (keine Haltestellen übersprungen, Rückfahrt ähnlich wie Hinfahrt)



Lösungsidee

- Relevante Kennzahlen zu jeder Reise sammeln (~20)
- Wie viel Zeit verging zwischen letzter Ticketkontrolle und Check-Out?
- Wie schnell bewegte sich der Kunde während dem Check-Out fort?
- Passen mehrere Reisen eines Kunden am gleichen Tag zusammen? (keine Haltestellen übersprungen, Rückfahrt ähnlich wie Hinfahrt)
- ...



Lösungsidee

- Relevante Kennzahlen zu jeder Reise sammeln (~20)
- Wie viel Zeit verging zwischen letzter Ticketkontrolle und Check-Out?
- Wie schnell bewegte sich der Kunde während dem Check-Out fort?
- Passen mehrere Reisen eines Kunden am gleichen Tag zusammen? (keine Haltestellen übersprungen, Rückfahrt ähnlich wie Hinfahrt)
- ...
- Viele der normalen Reisen werden ähnliche Kennzahlen aufweisen

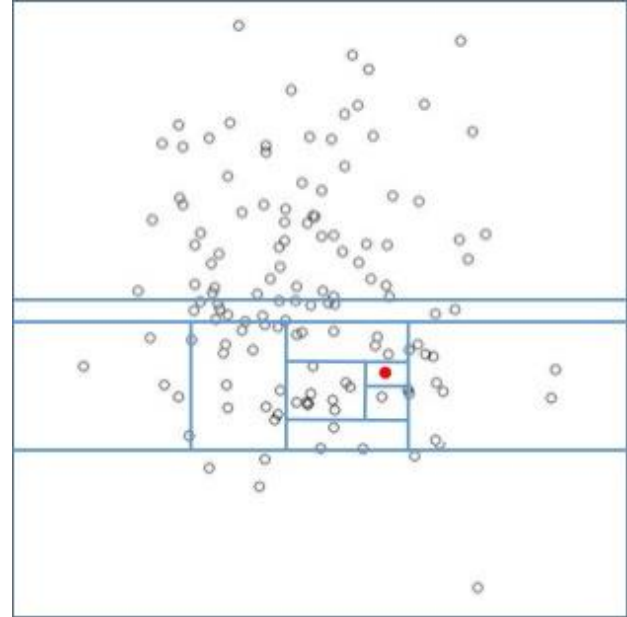


Lösungsidee

- Relevante Kennzahlen zu jeder Reise sammeln (~20)
- Wie viel Zeit verging zwischen letzter Ticketkontrolle und Check-Out?
- Wie schnell bewegte sich der Kunde während dem Check-Out fort?
- Passen mehrere Reisen eines Kunden am gleichen Tag zusammen? (keine Haltestellen übersprungen, Rückfahrt ähnlich wie Hinfahrt)
- ...
- Viele der normalen Reisen werden ähnliche Kennzahlen aufweisen
- Reisen mit auffälligen / ungewöhnlichen Kennzahl-Kombinationen automatisch erkennen (anomaly detection)



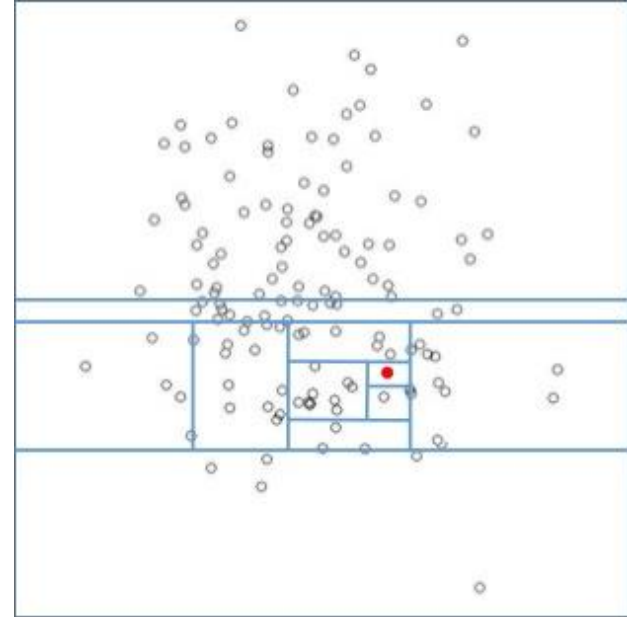
Isolation forest





Isolation forest

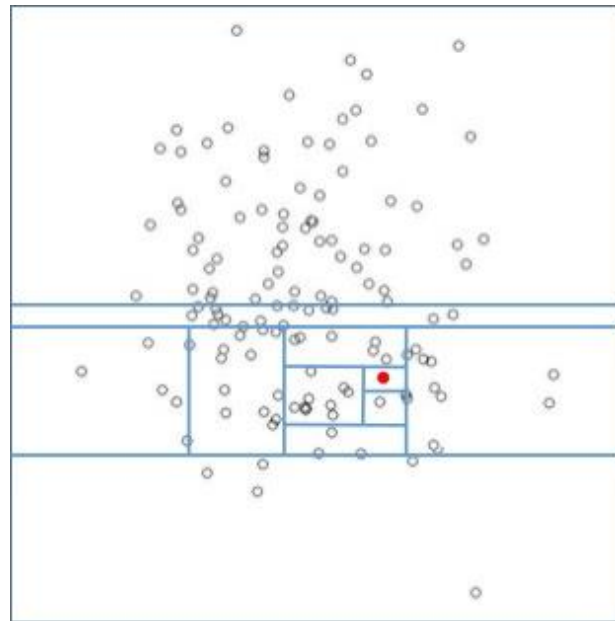
- Idee: Wenn man die Fläche (den Raum) zufällig unterteilt, werden isolierte Punkte schneller alleine dastehen





Isolation forest

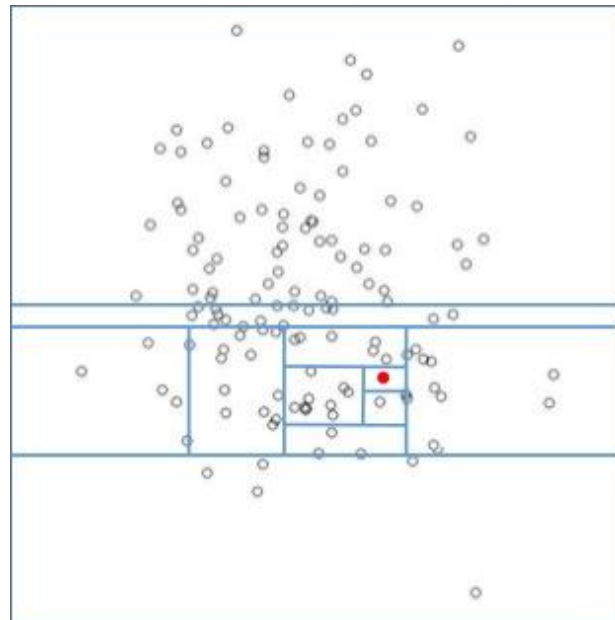
- Idee: Wenn man die Fläche (den Raum) zufällig unterteilt, werden isolierte Punkte schneller alleine dastehen
- Also ist die Anzahl Unterteilungen bis zur Isolation ein Indikator für „Normalität“





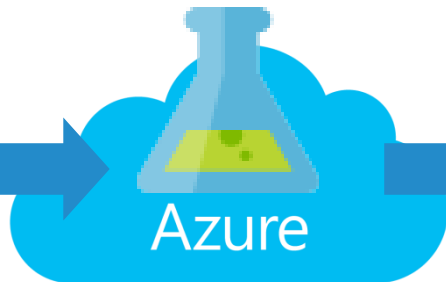
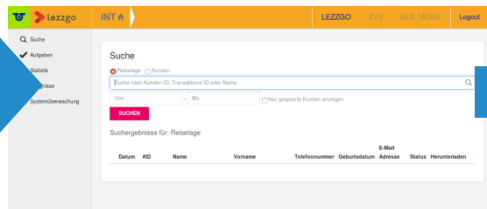
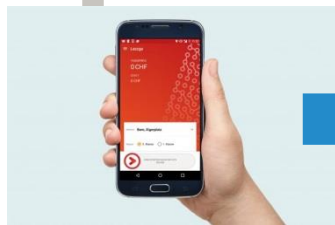
Isolation forest

- Idee: Wenn man die Fläche (den Raum) zufällig unterteilt, werden isolierte Punkte schneller alleine dastehen
- Also ist die Anzahl Unterteilungen bis zur Isolation ein Indikator für „Normalität“
- Durchschnitt vieler zufälliger Unterteilungen gibt eine zuverlässigere Bewertung





Pipeline



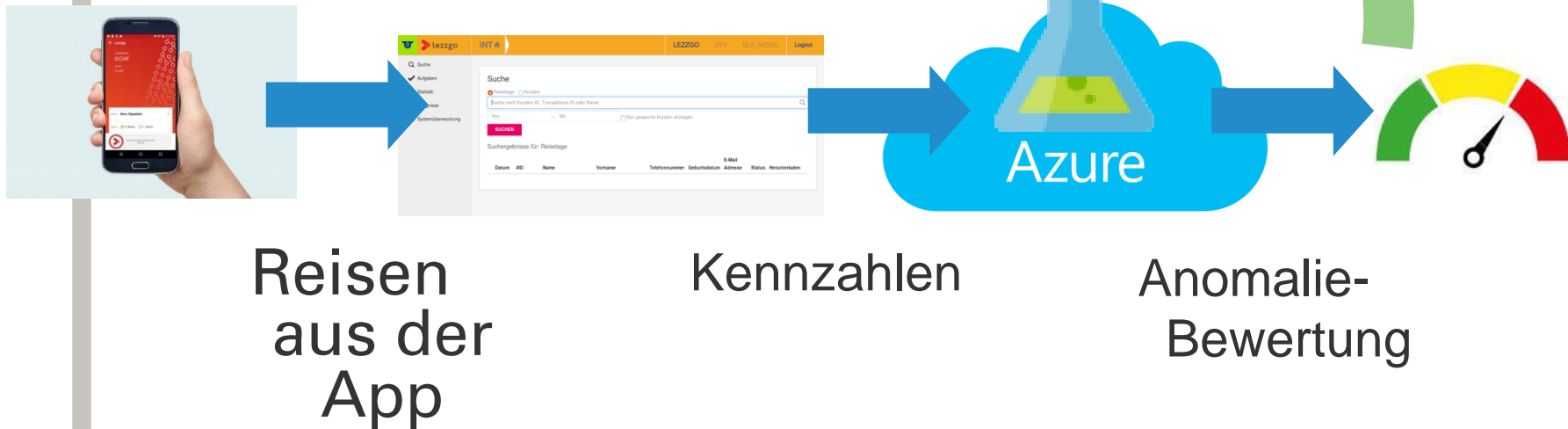
Reisen
aus der
App

Kennzahlen

Anomalie-
Bewertung

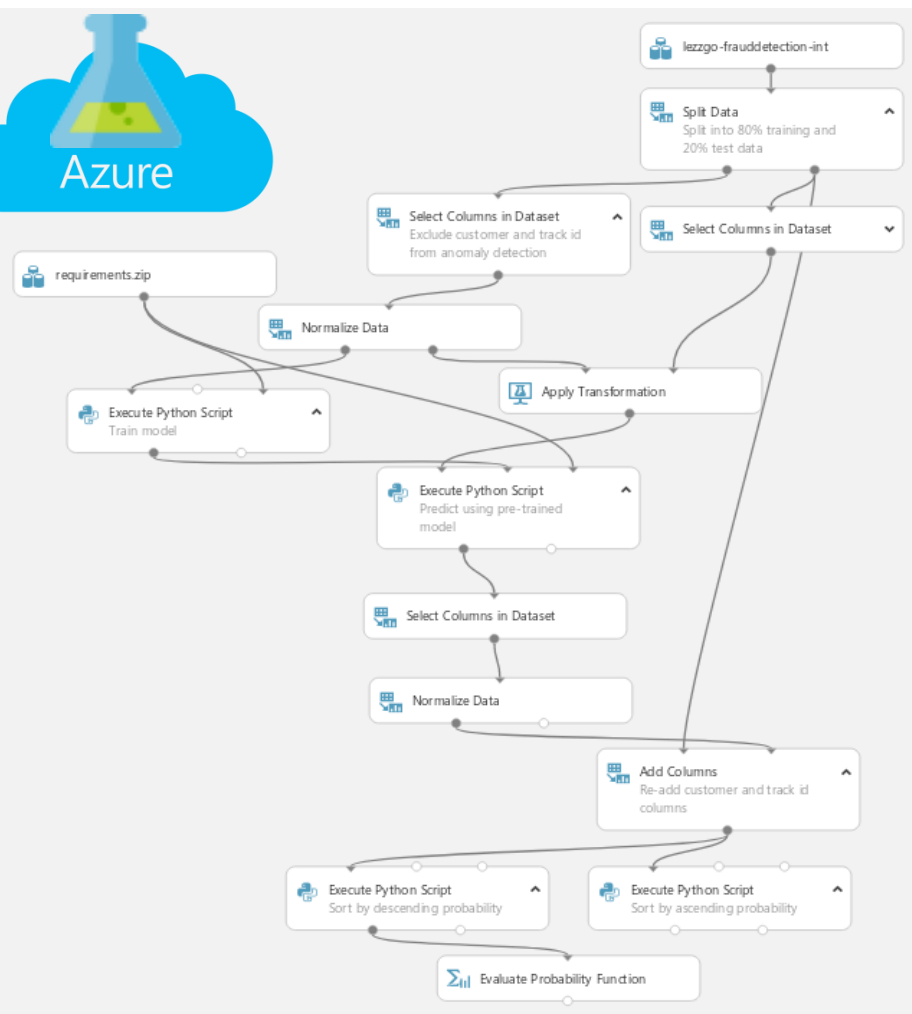
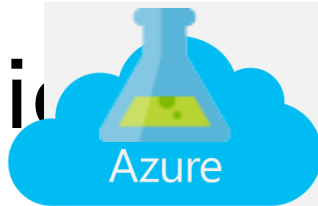


Pipeline





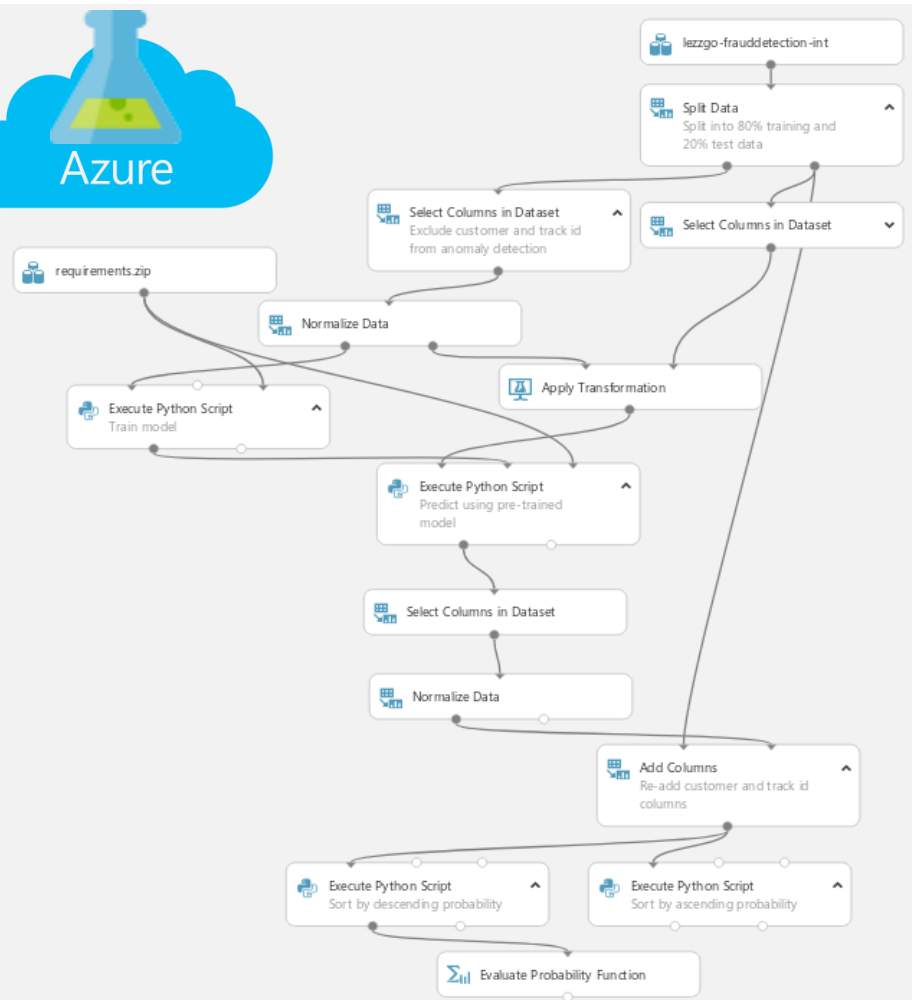
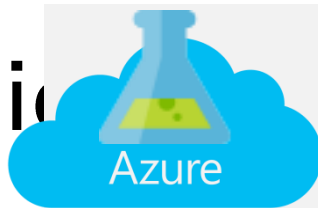
Azure ML Studio





Azure ML Studio

1. Daten-Pipelines per click-and-drag zusammenstellen

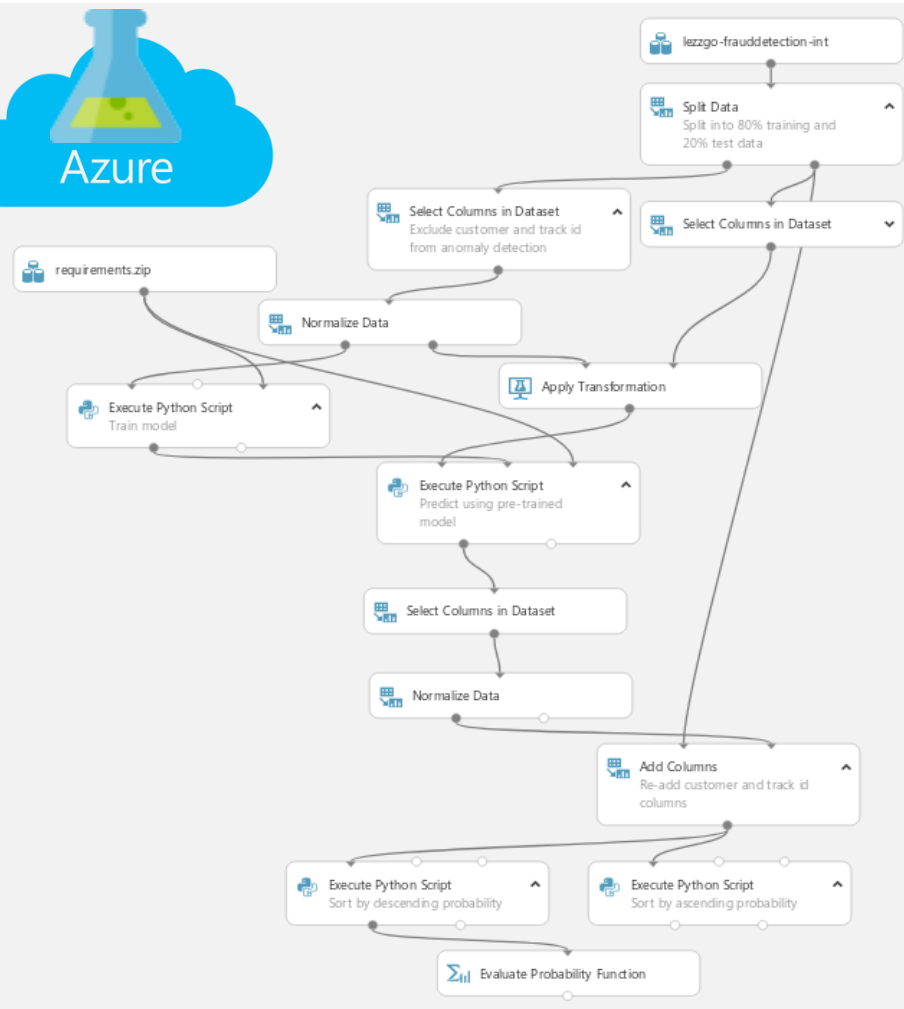
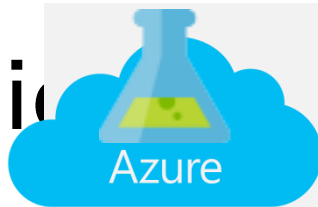




Azure ML Studio

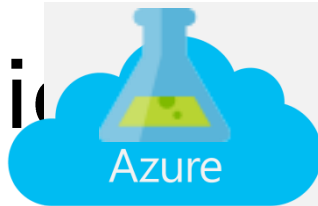
1. Daten-Pipelines per click-and-drag zusammenstellen

- Eingebaute Bausteine (Normalisierung, Daten aufteilen, einfache ML Algorithmen) oder custom Python Blöcke





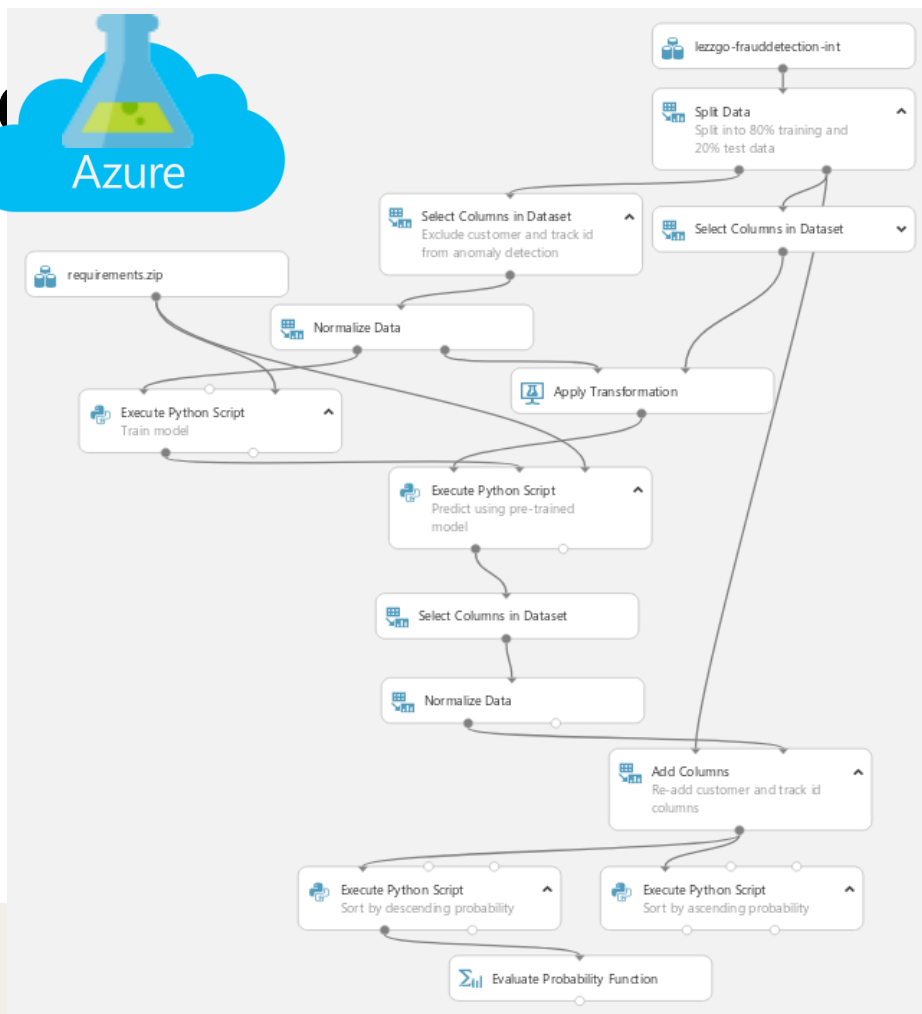
Azure ML Studio



1. Daten-Pipelines per click-and-drag zusammenstellen

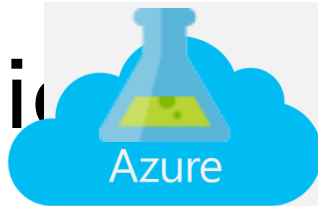
- Eingebaute Bausteine (Normalisierung, Daten aufteilen, einfache ML Algorithmen) oder custom Python Blöcke

2. Mit CSV-Daten trainieren





Azure ML Studio

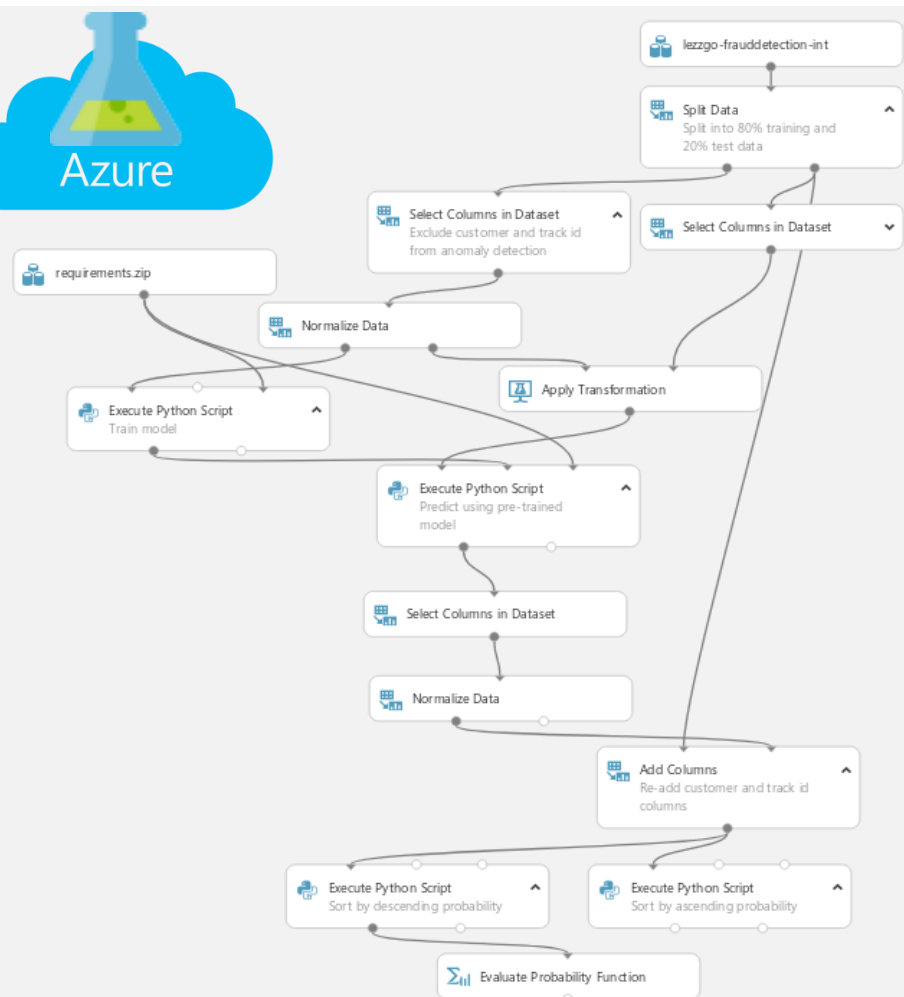


1. Daten-Pipelines per click-and-drag zusammenstellen

- Eingebaute Bausteine (Normalisierung, Daten aufteilen, einfache ML Algorithmen) oder custom Python Blöcke

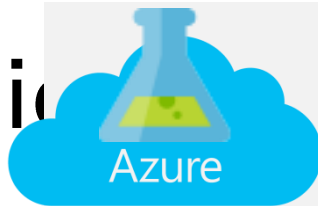
2. Mit CSV-Daten trainieren

3. Als Web Service veröffentlichen





Azure ML Studio



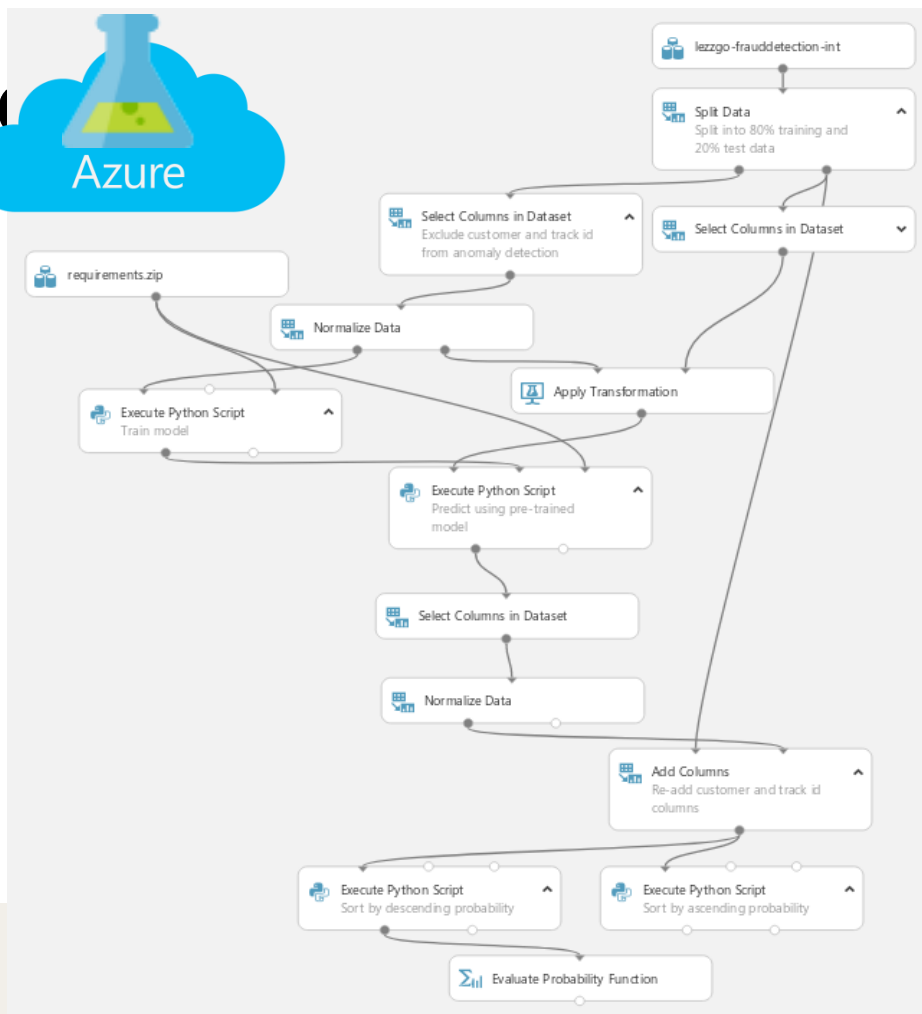
1. Daten-Pipelines per click-and-drag zusammenstellen

- Eingebaute Bausteine (Normalisierung, Daten aufteilen, einfache ML Algorithmen) oder custom Python Blöcke

2. Mit CSV-Daten trainieren

3. Als Web Service veröffentlichen

4. Anomalie-Bewertung für neue Reisen abfragen





Qualitätskontrolle





Qualitätskontrolle

- Testfahrer der BLS begehen absichtlich Missbrauch
- Optimieren der Parameter der Pipeline sodass die bekannten Missbrauchsfälle als möglichst auffällig bewertet werden
- Später alle bekannten Missbrauchsfälle vom Trainings-Datenset ausschliessen





Qualitätskontrolle

- Testfahrer der BLS begehen absichtlich Missbrauch





Qualitätskontrolle

- Testfahrer der BLS begehen absichtlich Missbrauch
- Optimieren der Parameter der Pipeline sodass die bekannten Missbrauchsfälle als möglichst auffällig bewertet werden





Qualitätskontrolle

- Testfahrer der BLS begehen absichtlich Missbrauch
- Optimieren der Parameter der Pipeline sodass die bekannten Missbrauchsfälle als möglichst auffällig bewertet werden
- Später alle bekannten Missbrauchsfälle vom Trainings-Datenset ausschliessen





Machine Learning

- Erklärbarkeit der Resultate
- Datenqualität ist wichtig
- Datenschutz ist aufwändig



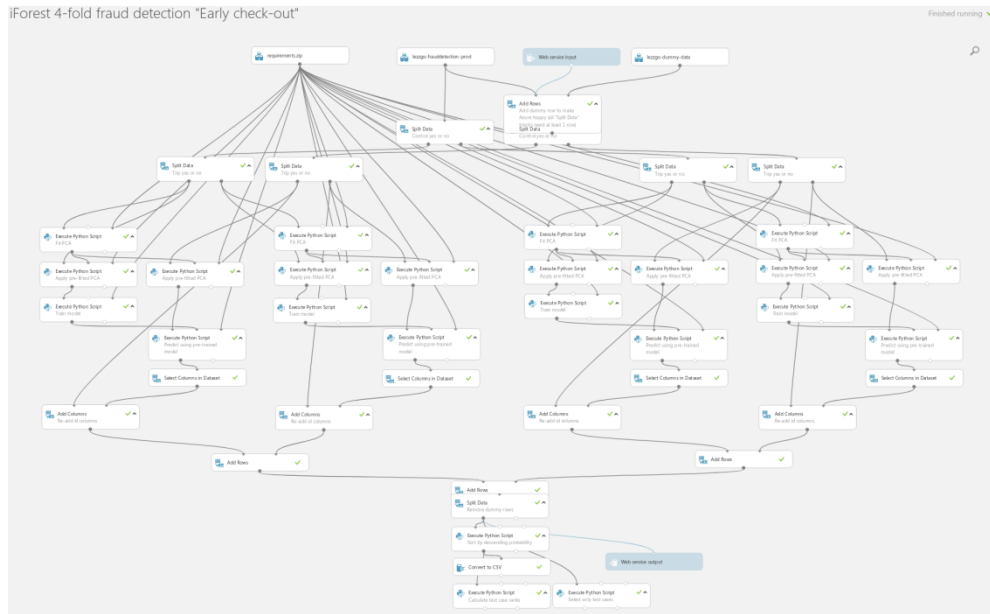


Azure ML Studio

- Gut für schnelle und flexible Experimente

- Gebaut für grosse Big Data Applikationen, Web Service klingt einfacher als es ist

- Als Betriebsplattform wegen fehlendem Versioning und Environment management noch nicht sehr ausgereift

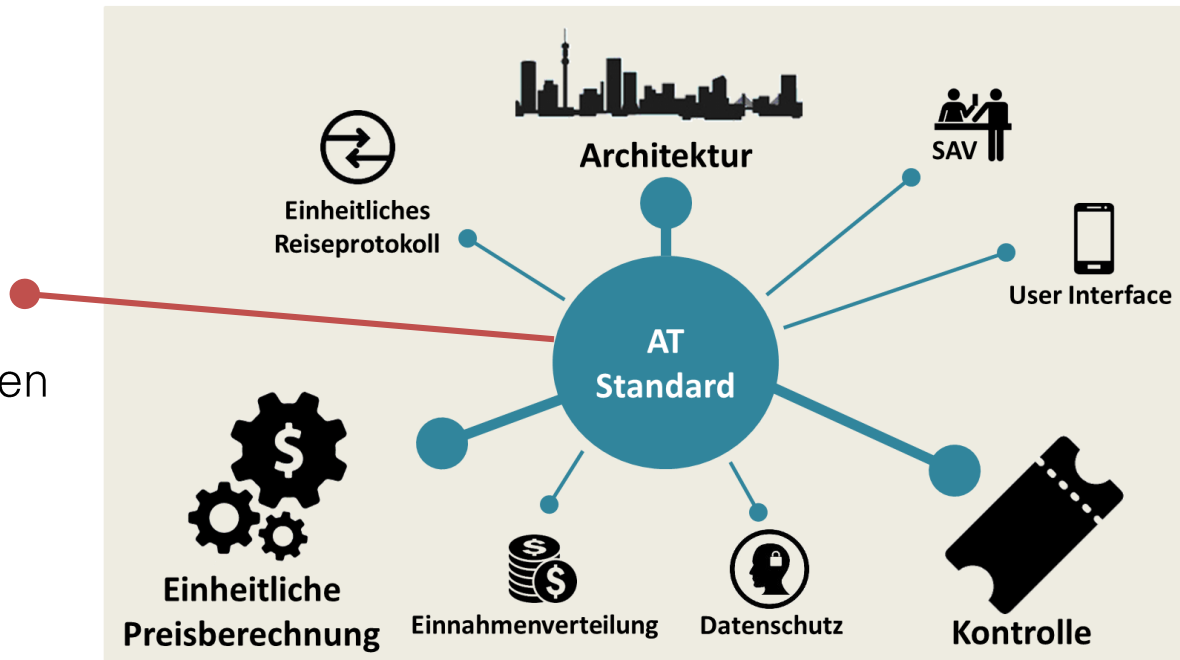




Ausblick



Missbrauch:
Neue Anforderungen
AT Standard &
Zertifizierung





Einsteigen, fahren, fahren, fahren,... bezahlen!