ZHAW School of Engineering  
Konzepte des maschinellen Lernens - KONZML

Elektrotechnik 2019

Projektarbeit zum Thema

**Crop Recommendation**

**with ML**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Barnes Toby - barnetob

Seminar: KONZML  
Fachsemester: Herbst 2021  
  
Abgabedatum: 10.01.2022

Dozent: Weinmann Thomas Oskar (weto)

# Einleitung

Technologien wie künstliche Intelligenz (KI), Robotik, das Internet der Dinge (IoT), Edge Computing, 5G und Blockchain haben alle das Potenzial, die Landwirtschaft effizienter, nachhaltiger und wettbewerbsfähiger zu machen. Die digitale Transformation der Landwirtschaft wird die Zusammenarbeit entlang der Wertschöpfungskette erleichtern, Landwirte unterstützen und Chancen für innovative KMUs bieten. Bei dieser Arbeit soll anhand eines [Datensatzes](https://www.kaggle.com/atharvaingle/crop-recommendation-dataset) ein Machine Learning Algorithmus trainiert werden, um ein Vorhersagemodell zu erstellen, dass auf der Grundlage verschiedener Parameter, die am besten geeigneten Pflanzen für den Anbau auf einem Ackerfeld empfiehlt. Die Eingabeparamater bestehen aus den nachfolgenden Definitionen:

* **N** - Verhältnis des Stickstoffgehalts im Boden:
* **P** - Verhältnis des Phosphorgehalts im Boden
* **K** - Verhältnis des Kaliumgehalts im Boden
* **temperature** - Temperatur in Grad Celsius
* **humidity** - relative Feuchtigkeit in %
* **ph** - ph-Wert des Bodens
* **rainfall** - Niederschlag in mm

Es werden dabei 22 Pflanzen prognostiziert. (Reis Mais Kichererbse Kidneybohnen Taubenbohnen Mottenbohnen, Mungbohne, Urdbohne, Linse, Granatapfel, Banane, Mango, Weintrauben, Wassermelone, Zuckermelone, Apfel, Orange, Papaya, Kokosnuss, Baumwolle, Jute, Kaffee)

# ML – Pipeline

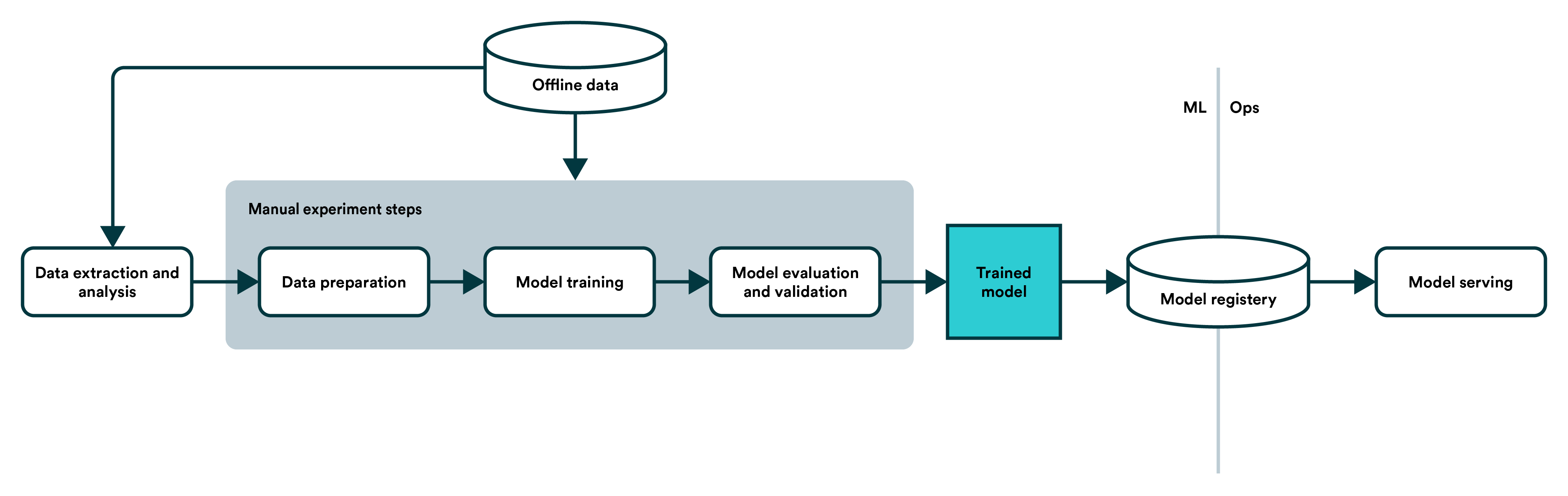


Abbildung 1 - Pipline von einem ML-Model

# Vorverarbeitungsschritte

## Vorevaluierung

### Scatter Plots

Pair plot erwähnen

|  |  |
| --- | --- |
| Abbildung 2 - 2D-Scatter Plot | Abbildung 3 - 3D-Scatter Plot |
|  |  |

### Korrelationmatrix

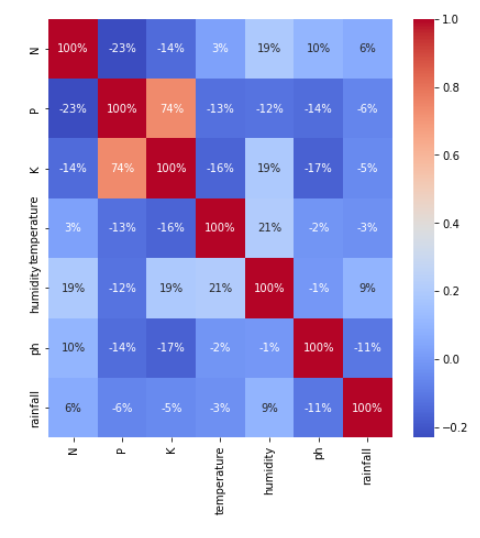
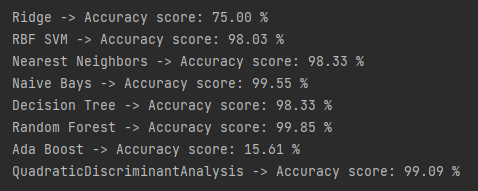
Eine Korrelationsmatrix ist eine Tabelle mit Korrelationskoeffizienten zwischen Variablen. Jede Zelle in der Tabelle zeigt die Korrelation zwischen zwei Variablen an. Der Sinn und Zweck dieser Matrix ist es, dass Muster in den Daten erkannt werden. Beim verwendeten Datensatz wurde in der generierten Korrelationsmatrix (siehe Abbildung 4) erkannt, dass der Phosphorgehalt und der Kaliumgehalt im Boden stark korrelieren. Jedoch gibt es kaum Korrelation zwischen den anderen Parametern.

Abbildung 4 - Korrelationmatrix

## Modellselektion



# Umsetzung

## ML – Model

### Gaussian Naive Bays - Algorithmus

### Random Forest - Algorithmus

## Evaluation mit Confusion Matrix

|  |  |
| --- | --- |
| Abbildung 5 - Gaussian NB Evaluation | Abbildung 6 - Random Forest Evaluation |
|  |  |

## Random Forest Decision Tree

# Fazit & Diskussion

# Quellenverzeichnis