

„Prädiktive Analyse von Weinproduktionsdaten Österreichs zur Vorhersage von Produktionsmengen einzelner Betriebe“

In dieser Diploma-Arbeit wurden Daten des Bundesministeriums für Land- und Forstwirtschaft, Klima- und Umweltschutz, Regionen und Wasserwirtschaft (BMLUK) analysiert, mit dem Ziel, ein statistisches Modell zu entwickeln, das die Erntemeldungen der österreichischen Weinbaubetriebe für das Folgejahr prognostizieren kann. Die erhobenen Daten stammen größtenteils aus den jährlichen Erntemeldungen und den Kontrollen der Bundeskellereiinspektion (BKI).

Die Motivation der Arbeit lag auf der Sammlung und Analyse von Daten, um die produzierten Weinmengen und die Anbaufläche der Winzerbetriebe nach Regionen über mehrere Jahre hinweg zu untersuchen. Das daraus resultierende Modell soll es ermöglichen, die prognostizierten Mengen für die kommende Weinperiode vorherzusagen.

Die Zielsetzung der Arbeit war es, ein statistisches Modell zu erstellen, welches die Produktionsmenge der einzelnen Winzer für die nächste Meldeperiode prognostiziert. Die Bundeskellereiinspektor:innen könnten dann in der neuen Kontrollperiode (neues Weinjahr) gezielt jene Winzer verstärkt kontrollieren, bei denen die Meldung der Weinbaubetriebe stark vom prognostizierten Wert für diesen Betrieb abweicht. Die Erwartung ist, dass Kontrollen dadurch eine deutliche Effizienzsteigerung erlangen werden.

Als Datenquellen standen die Daten der Weinbaubetriebe von der BKI aus den Jahren 2019 bis 2023 zur Verfügung. Weiters wurden monatliche Wetterdaten von der Webseite der GeoSphere Austria für denselben Zeitraum heruntergeladen. Dafür stehen für 1.134 Messstationen in ganz Österreich diverse Parameter zur Verfügung. Wobei 33 Parameter als interessant eingestuft wurden und für alle relevanten Messstationen in den vier Bundesländern Burgenland, Niederösterreich, Steiermark und Wien erhoben wurden (Anmerkung: die Betriebe aus der Weinbauregion Bergland – insgesamt weniger als 1 % der Betriebe – wurden nicht berücksichtigt).

In den Daten der Weinbaubetriebe ist lediglich die Gemeindekennzahl (GKZ) als regionaler Hinweis enthalten. Um hier den entsprechenden Gemeindenamen zu erhalten, damit die Daten mit den Wetterdaten verknüpfen werden können, wurde das Gemeindeverzeichnis, welches auf der Webseite der Statistik Austria verfügbar ist, verwendet.

Im Zuge der Datenaufbereitung wurde zunächst eine Datenbereinigung durch Entfernung irrelevanter und fehlerhafter Daten vorgenommen. Danach erfolgten diverse Datennormalisierungen und Datentransformationen, bevor eine Zuordnung der Wetterdaten zu den Weinbaugemeinden durchgeführt wurde. Einzelne fehlende Wetterdaten wurden anschließend durch Imputation aufgefüllt.

Für die Daten aus den Erntemeldungen und die weiteren extrahierten Daten wurden nach der Datenqualitätsprüfung die Datenanalysen erstellt. Als statistische Methoden wurden zunächst Zeitreihenanalysen und deskriptive Analysen zur Untersuchung der Daten angewandt. Mit Hilfe der deskriptiven Analysen gelingt es Größenordnungen, Verteilungen und Streuungen der relevanten Merkmale zu prüfen. Mit Hilfe der Zeitreihenanalysen werden Vergleiche der Erntemengen und der Wetterdaten über die Jahre angestellt.

Nachdem man sich mit den ersten statistischen Analysen einen Überblick über die verfügbaren Daten verschafft hat, kann mit der Modellerstellung begonnen werden. Als abhängiges Merkmal (zu erklärende Variable) wird die Produktionsmenge der einzelnen Weinbaubetriebe über die fünf Beobachtungsjahre herangezogen. Als unabhängige Merkmale (erklärende Variablen) dient einerseits die gemeldete Weinbaufläche und andererseits diverse Parameter aus den Wetterdaten. Für die Modellentwicklung fanden multivariate Regressionsmodelle, generalisierter linearer Modelle (GLM) und Methoden des maschinellen Lernens, ihre Anwendung, um ein optimales Prognosemodell zu entwickeln. Die prädiktive Modellbildung ermittelt die wichtigsten Prädiktoren aus den unabhängigen Merkmalen, um die Produktionsmenge möglichst gut zu prognostizieren. Als das geeignetste Modell kristallisierte sich ein Random Forest-Modell heraus.

Die entwickelte Methode zur Prognose der Erntemengen bieten der Bundeskellereiinspektion die Möglichkeit, gezieltere Kontrollen durchzuführen. Die Wetterdaten und Anbauflächen liefern wertvolle Einblicke in Faktoren, welche die Produktionsmenge beeinflussen. Eine mögliche zukünftige Verfeinerung der Modelle könnte die Einbeziehung weiterer Wetterparameter und die Ausweitung zu Modellen auf Bundeslandebene umfassen.

Außerdem wäre es mit detaillierteren Daten der Weinbaubetriebe möglich, gesonderte Modelle nach den produzierten Mengen in den Bereichen Landwein und Qualitätswein und nach den produzierten Mengen im Weißwein- und Rotwein-Bereich zu erstellen. Es ist damit zu rechnen, dass die Genauigkeit und die Güte der Modelle zunehmen würde.

Mit dieser Diploma-Arbeit wurde ein erster Schritt unternommen, um mittels Datenanalysen die Effizienz der Bundeskellereiinspektion zu verbessern und dabei einen noch umfassenderen Blick auf die Weinwirtschaft in Österreich zu erlangen.