

# Richtlinie zum „Software-Sprint“

[SAT]

---

## *Schlussbericht*

Zuwendungsempfänger:

[Sherif Elsayed]

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS24S32 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

### **Kurze Darstellung der Aufgabenstellung und Motivation**

Die Motivation hinter dem Projekt war es, Studierende und Forschende dabei zu unterstützen, den Umgang mit Satellitendaten zu erlernen, um Machine-Learning-Modelle zu entwickeln. Satellitendaten spielen mittlerweile eine entscheidende Rolle im Verständnis und der Bekämpfung des Klimawandels sowie bei der Förderung nachhaltiger Praktiken. Das Ziel von @Sat ist es, die Generierung von KI-Modellen in der Geoinformatik zu beschleunigen, indem die Erstellung von Trainingsdatensätzen vereinfacht wird. Dadurch wird die breite Anwendung öffentlich finanzierter Satellitendaten im Alltag ermöglicht. Obwohl sowohl die Zugänglichkeit von Satellitendaten als auch die Entwicklung von KI-Modellen erhebliche Fortschritte gemacht haben, bleibt die Vorbereitung von Trainingsdatensätzen eine zeitaufwendige Herausforderung für Geoinformatiker\*innen. @Sat bietet die notwendige Effizienz und Flexibilität, um mit Satellitendaten und Datensätzen wie EuroCrops auf noch nie dagewesene Weise KI-Modelle zu erstellen. Diese Modelle können zur Analyse verschiedenster gesellschaftlicher und umweltbezogener Fragestellungen eingesetzt werden und tragen so zu datenbasierten Lösungen für Nachhaltigkeit und Klimaschutz bei.

Der Ansatz begann mit einer Untersuchung verschiedener Satellitendatenanbieter wie Google Earth Engine und Sentinel Hub. Dies war ein wichtiger erster Schritt, da er den gesamten Arbeitsablauf des Projekts beeinflusst. Die ersten zwei Monate wurden genutzt, um beide Plattformen hinsichtlich Kosten, Zuverlässigkeit und ihrer Eignung für die Projektziele zu evaluieren. Anschließend verlagerte sich der Fokus auf die Arbeit mit den Satellitendaten selbst, da diese in der Regel umfangreiche Expertise in Bildverarbeitung, Fernerkundung, Cloud Computing und Machine Learning erfordern. Dabei gab es zahlreiche Versuche und Fehler, um den Arbeitsablauf zu optimieren und ihn so reibungslos wie möglich zu gestalten.

Ein weiterer wichtiger Schritt war die Auswahl der geeigneten Satellitenmissionen, da tausende Missionen die Erde umkreisen. Unser Fokus lag hauptsächlich auf Satellitenmissionen, die in Deutschland entwickelt wurden, wie Sentinel-2 und Sentinel-1. Diese bieten den Vorteil, dass ihre Daten kostenlos verfügbar sind, was ideal zu den Projektzielen passt. Abschließend wurden deutsche Trainingsdatensätze für die Klassifikation von Nutzpflanzen verwendet, um den gesamten Workflow des Projekts zu demonstrieren.

## **Beitrag des Projektes zu den Zielen der Förderinitiative „Software-Sprint“**

Die ursprüngliche Zielgruppe des @Sat-Projekts waren Forschende und Studierende, die erheblich davon profitiert haben, Trainingsdatensätze mit nur wenigen einfachen Schritten zu erstellen. Dies reduziert die benötigte Zeit für die Entwicklung satellitengestützter Klimaschutzlösungen erheblich. Das Projekt steht im Einklang mit den Zielen der "Software-Sprint"-Initiative, indem es die Erstellung von Trainingsdatensätzen für KI-Modelle in der Geoinformatik vereinfacht und somit die breitere Anwendung öffentlich finanzierter Satellitendaten im Alltag ermöglicht. Durch die Nutzung kostenloser Satellitendaten und frei verfügbarer, von deutschen Institutionen bereitgestellter Beschriftungsdatensätze bietet @Sat eine zugängliche und effiziente Möglichkeit, KI in die geospatiale Analyse zu integrieren. Indem es die notwendige Effizienz und Flexibilität bietet, um KI-Modelle mit Satellitendaten und Datensätzen wie EuroCrops zu erstellen, ermöglicht @Sat die Analyse verschiedener gesellschaftlicher und umweltbezogener Fragestellungen. Dies trägt zu datenbasierten Lösungen für Nachhaltigkeit und Klimaschutz bei und unterstützt direkt die Ziele der "Software-Sprint"-Initiative, innovative, Open-Source-Projekte zu fördern, die drängende gesellschaftliche Herausforderungen adressieren.

Zusätzlich wurde eine weitere kostenlose Lösung entwickelt, die von Landwirt\*innen genutzt werden kann, um ihr Land zu analysieren, die Erntequalität zu bewerten und den Einsatz von Ressourcen wie Wasser und Dünger zu optimieren, was zu nachhaltigeren landwirtschaftlichen Praktiken beiträgt.

## **Ausführliche Darstellung der Ergebnisse**

Das Ziel dieses Projekts war die Entwicklung eines hochgradig automatisierten End-to-End-Arbeitsablaufs zum Herunterladen von Satellitendaten und zur Erstellung von Trainingsdatensätzen für Machine-Learning-Anwendungen, mit einem speziellen Fokus auf die Pflanzenklassifikation. Die endgültige Implementierung automatisierte erfolgreich etwa 90 % des gesamten Prozesses, was die manuelle Arbeitsbelastung erheblich reduzierte und die großflächige Datenvorbereitung effizienter und reproduzierbarer machte.

Der Arbeitsablauf beginnt mit der Suche und dem Herunterladen von Satellitenbildern aus zwei verschiedenen Erdbeobachtungsmissionen: Sentinel-1 und Sentinel-2. Sentinel-2 liefert

multispektrale optische Bilder, während Sentinel-1 Radardaten bereitstellt. Diese Datensätze sind komplementär und wertvoll für landwirtschaftliche Anwendungen, insbesondere in Regionen, in denen die Wolkenbedeckung häufig die Nutzbarkeit von optischen Daten allein einschränkt.

Nach dem ersten Herunterladen werden die Bilder einer Vorverarbeitung unterzogen, um sicherzustellen, dass alle Datenkanäle räumlich ausgerichtet und auf eine konsistente Auflösung neu abgetastet werden. In diesem Fall arbeiteten wir mit insgesamt 12 Kanälen—10 von Sentinel-2 und 2 von Sentinel-1. Der Ausrichtungsschritt stellt sicher, dass alle Kanäle die gleiche Geometrie und Auflösung teilen, was eine entscheidende Voraussetzung für die Kombination der Daten zu einem einheitlichen Datensatz ist, der für das Training von Segmentierungsmodellen geeignet ist.

Sobald die Satellitenbilder vorverarbeitet sind, werden sie weiter mit den Label-Daten abgeglichen. In diesem Projekt bestanden die Label-Daten aus einer Pflanzenklassifikationskarte, die als Ground-Truth für das überwachte Lernen diente. Die perfekte Ausrichtung zwischen Satellitenbildern und Klassifikationslabels—sowohl geometrisch als auch räumlich—war von wesentlicher Bedeutung für die genaue Schulung und Validierung von Machine-Learning-Modellen.

Nach der Ausrichtung wird der kombinierte Datensatz in kleinere Kacheln unterteilt. Diese Kacheln werden dann im NumPy-Format exportiert, einem Standardformat in Machine-Learning-Pipelines, das eine einfache Integration mit beliebigen Deep-Learning-Frameworks wie TensorFlow oder PyTorch ermöglicht. Dieser Kachelungsprozess hilft auch bei der Verwaltung der Speichernutzung und ermöglicht das parallele Training mit großen Datensätzen.

Anschließend haben wir ein KI-Modell zur Pflanzenklassifikation unter Verwendung der exportierten Daten erstellt, das eine Genauigkeit von 90 % erreichte. Dieser Schritt war entscheidend, um die Qualität der Ausgabedaten zu gewährleisten und die Effektivität des gesamten Arbeitsablaufs bei der Erstellung zuverlässiger und genauer Ergebnisse zu validieren.

Die einzigen verbleibenden manuellen Schritte im Arbeitsablauf betreffen die Überprüfung und den Ausschluss von Bildbereichen, die Fehler enthalten könnten—wie etwa wolkenbedeckte Regionen, Schatten oder Artefakte, die durch Sensorprobleme verursacht werden. Diese Entscheidung war absichtlich. Obwohl es technisch möglich ist, die Wolkenerkennung und Maskierung zu automatisieren, wurde dieser Schritt manuell belassen, um den Nutzern mehr Kontrolle und Flexibilität bei der Datenauswahl zu geben. In vielen realen Szenarien kann eine vollständig automatisierte Filterung zusätzliches Rauschen oder Fehldetektionen einführen, die letztendlich mehr Zeit und Aufwand erfordern, um sie zu korrigieren. Indem den Nutzern diese Entscheidung überlassen wird, bleibt der Arbeitsablauf robust, transparent und anpassungsfähig an spezifische Projektanforderungen.

Zusammenfassend lässt sich sagen, dass das entwickelte System eine optimierte und effiziente Methode bietet, um rohe Satellitendaten in qualitativ hochwertige Trainingsdatensätze für Pflanzenklassifikationsaufgaben umzuwandeln. Während der Kernarbeitsablauf größtenteils automatisiert ist, stellt die bewusste Aufnahme minimaler manueller Überprüfungsschritte sicher, dass der Arbeitsablauf die Flexibilität und Zuverlässigkeit bietet, die für vielfältige und dynamische landwirtschaftliche Umgebungen erforderlich sind.

Ergebnisdaten:



Wie Sie sehen können, ist das erste Bild für den Sentinel-2-Satelliten und es hatte eine Auflösung von 20 Metern. Jetzt ist es mit dem zweiten Bild ausgerichtet, das ein Sentinel-1-Bild ist, das von einem Radarinstrument aufgenommen wurde. Jetzt sind beide Satellitenbilder geometrisch korrigiert und neu abgetastet, um mit den Pflanzenklassifikations-Labels übereinzustimmen, die das dritte Bild darstellen.

Neben dem technischen Arbeitsablauf wurde auch eine funktionierende Plattform geliefert, die landwirtschaftlichen Akteuren—insbesondere Landwirten—ein praktisches Tool zur Analyse ihrer Felder bietet. Die Plattform unterstützt die Interpretation satellitenbasierter Erkenntnisse und zeigt die praktische Anwendbarkeit im Fernerkundungstechnologien in der Präzisionslandwirtschaft.



## **Zielgruppe, Nutzen und mögliche Weiterentwicklungen**

Die Vorteile für die Zielgruppe sind eindeutig: Durch die Vereinfachung des Prozesses zur Erstellung von Trainingsdatensätzen wird erhebliche Zeit gespart, sodass sich die Nutzer mehr auf die maschinellen Lernanwendungen konzentrieren können. Dies behebt einen wichtigen Engpass in diesem Bereich, da Daten immer der wichtigste Aspekt beim Aufbau einer maschinellen Lernanwendung bleiben werden. Zusätzlich haben Landwirt\*innen jetzt Zugang zu einer kostenlosen Plattform, mit der sie ihre Flächen einfach analysieren können, ohne für kommerzielle Plattformen bezahlen zu müssen.

Es gibt viele Ideen für die Weiterentwicklung des Projekts, wie zum Beispiel das Hinzufügen weiterer Satellitendatenmissionen und die Erweiterung der Vielfalt der Trainingsdatensätze. Der Fokus liegt derzeit hauptsächlich auf Segmentierungsproblemen wie der Klassifikation von Nutzpflanzen, aber es besteht Potenzial, auch andere Arten von Datensätzen und maschinellen Lernanwendungen zu erforschen.

## **Hat die Arbeit in dem Projekt Dich in Deiner persönlichen, fachlichen Weiterentwicklung unterstützt?**

Definitiv, ich bin dankbar für diese Gelegenheit, da sie mir geholfen hat, meine praktischen Fähigkeiten in verschiedenen Bereichen wie maschinellem Lernen, Fernerkundung und Cloud Computing weiterzuentwickeln. Ich habe diese Themen in meinem Masterstudium gelernt, das ich 2014 an der Technischen Universität München abgeschlossen habe. Seitdem hatte ich keine Gelegenheit, das Gelernte anzuwenden. Dieses Projekt war der beste Neustart, da ich mich intensiv mit all diesen Bereichen beschäftigen konnte, besonders da sich Software- und KI-Technologien in rasantem Tempo weiterentwickeln. Es war die perfekte Gelegenheit, mich mit den aktuellen Technologietrends vertraut zu machen und meine Fähigkeiten weiter auszubauen.

## **Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben**

Das Ziel des Projekts war es, den gesamten Arbeitsablauf für das Herunterladen von Satellitendaten und die Erstellung von Trainingsdatensätzen für Machine-Learning-Anwendungen vollständig zu automatisieren. Das Projektergebnis konnte 90 % des Arbeitsablaufs automatisieren, wobei nur wenige manuelle Schritte noch erforderlich waren, wie zum Beispiel das Ausschließen von korrupten Bereichen aus der Bildszene.

Allerdings würde die Erreichung eines höheren Automatisierungsgrades mehr Zeit erfordern, da das Thema sehr komplex ist und viele verschiedene Ingenieurdisziplinen miteinander verknüpft sind.

Darüber hinaus hatte das Projekt das Ziel, Datensätze für verschiedene Machine-Learning-Probleme zu erstellen, wie beispielsweise Bildsegmentierung und Objekterkennung. Am Ende lag der Fokus jedoch ausschließlich auf der Segmentierung. Ich habe den Arbeitsablauf so weit wie möglich automatisiert und die Komplexität reduziert, indem ich mich darauf konzentrierte, einen End-to-End-Arbeitsablauf für ein spezifisches Problem zu bieten: die Klassifizierung von Nutzpflanzen mithilfe von Bildsegmentierungstechniken.

Zusätzlich haben wir auch eine funktionierende Plattform für die Landwirtschaft bereitgestellt, die von Landwirten genutzt werden kann, um ihre Felder zu analysieren.

## **Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer**

<https://github.com/sherifelsayed1/SAT-PrototypeFund>

## **Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung**

das Projekt hatte ursprünglich das Ziel, Daten über Sentinel Hub herunterzuladen, aber nach einiger Zeit entdeckten wir eine Alternative, nämlich Google Earth Engine. Daher verlagerte sich der Fokus darauf, beide Plattformen parallel zu erlernen, um ihre Fähigkeiten und Zuverlässigkeit zu bewerten, was die Komplexität des Projekts erhöhte. Am Ende entschieden wir uns, Google Earth Engine zu bevorzugen, da es besser mit den Zielen des Projekts in Einklang steht. Es ist kostenlos und bietet mehr Rechenfunktionen im Vergleich zu Sentinel Hub, das ein Abonnement erfordert.

Zusätzlich stellten wir fest, dass viele Landwirtinnen *aufgrund der Auswirkungen des Klimawandels mit ihren Flächen kämpfen*. Daher konzentrierten wir uns nicht nur darauf, Lösungen für Forscherinnen und Studierende zu entwickeln, sondern auch Landwirt\*innen zu unterstützen, indem wir eine vollständig kostenlose Plattform zur Flächenanalyse entwickelten. Diese Entscheidung erhöhte den Arbeitsaufwand, machte jedoch auch das Projektergebnis wirkungsvoller und half verschiedenen gesellschaftlichen Gruppen.

## **Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen**

Es gab keine Entwicklungen von anderen Personen oder Institutionen, aber ich habe viel durch Gespräche mit Forschern und Studenten vor Ort gelernt, um ihre Bedürfnisse besser zu verstehen. Außerdem sprach ich mit normalen Menschen bei Veranstaltungen, um die Bedürfnisse der Gesellschaft zu verstehen, und nicht nur aus theoretischer Forschung. Diese Erfahrungen haben mir geholfen, den Entwicklungsansatz zu überdenken und die Art und Weise, wie ich die finalen Ergebnisse präsentiere, zu ändern.

Zum Beispiel bevorzugten es die meisten Studenten und Forscher, Python-Notebooks zu haben, die mit nur wenigen einfachen Klicks leicht zu bedienen sind, anstatt komplexe Bibliotheken zu verwenden, die viel Zeit für die Installation und das Erlernen erfordern.

Ein weiteres Beispiel ist, dass viele Landwirte wirklich nach Satellitenlösungen suchen, um ihre Farmen zu überwachen, aber sie beklagten sich über die hohen Kosten bestehender Lösungen und das mangelnde Vertrauen in diese Plattformen. Daher habe ich eine kostenlose Plattform entwickelt, die sie nutzen können, um zu verstehen, wie diese Technologie funktioniert. Auf diese Weise können sie, wenn sie sich für kommerzielle Software entscheiden, dies tun, nachdem sie die kostenlose Plattform genutzt haben, um ihre Bedürfnisse besser zu verstehen, bevor sie für solche Plattformen bezahlen müssen.