



# SafeCycling

**Prognose des Gefahrenpotenzials  
im Berliner Radverkehr**

Modul Machine Learning bei  
Prof. Dr. Helena Mihaljević

Nicole Driebe | Sabine Matthies | Lisa Rübel



**Hochschule für Technik  
und Wirtschaft Berlin**

University of Applied Sciences



## Inhalt

1. Einleitung
2. Beschreibung der verwendeten Daten
  - 2.1 SimRa-Datensatz
  - 2.2 OSM-Datensatz
  - 2.3 Integration der Datensätze
3. Vorgehensweise
  - Sprint 1 - Verständnis für die Aufgabenstellung und Recherche von Datensätzen
  - Sprint 2 - Explorieren der Datensätze
  - Sprint 3 - Explorieren der OSM-Daten und Bereinigung unserer Datensätze
    - OSM-Daten
    - SimRa-Daten
  - Sprint 4 - Verknüpfung der Datensätze
  - Sprint 5 - Modelltraining
  - Sprint 6 - Evaluation und Darstellung der Ergebnisse
4. Reflexion
  - Erreichte Ziele
  - Ausblick: Nächste Schritte



## 1. Einleitung

Als begeisterte Radfahrerinnen erleben wir täglich die Herausforderungen, die das Fahrradfahren im Berliner Straßenverkehr mit sich bringt. Trotz des wachsenden Trends hin zu umweltfreundlicheren Verkehrsmitteln und einer verstärkten Nutzung des Fahrrads als Transportmittel bleibt das Risiko von Unfällen, insbesondere durch den Autoverkehr, ein großes Problem. Angesichts dessen wird die Frage, wie das Gefahrenpotenzial für Radfahrende präzise prognostiziert und gemindert werden kann, immer relevanter.

Das Projekt **SafeCycling** zielt darauf ab, durch den Einsatz von Machine-Learning-Methoden eine fundierte Prognose des Gefahrenpotenzials im Berliner Radverkehr zu entwickeln. Als Datengrundlage wird der **SimRa-Datensatz** verwendet, der durch Crowdsourcing Informationen über gefährliche Situationen im Radverkehr sammelt. Ergänzend hierzu wird der **OpenStreetMap (OSM)-Datensatz** herangezogen, um infrastrukturelle Merkmale wie zum Beispiel den Straßenbelag und die zulässige Höchstgeschwindigkeit zu berücksichtigen.

Eine zentrale Frage in unserem Projekt lautet: **Wie gut lässt sich das Gefahrenpotenzial für Radfahrende in Berlin auf Basis der genannten Merkmalen und maschinellen Modellen vorhersagen?**

## 2. Beschreibung der verwendeten Daten

### 2.1 SimRa-Datensatz

Der SimRa-Datensatz (erhoben im Rahmen des Pilotprojekts "Sicherheit im Rad-verkehr") umfasst detaillierte Informationen zu Beinaheunfällen, die von Radfahrer\*innen über eine Smartphone-App aufgezeichnet und gemeldet werden. Die App erfasst GPS-Daten der Fahrtrouten und nutzt Beschleunigungssensoren, um Gefahrensituationen wie plötzlichem Bremsen oder Ausweichen zu erkennen. Nach der Fahrt können die Nutzer\*innen diese Situationen kategorisieren, ergänzen und auf die Projektserver hochladen.

- **Erhebungszeitraum:** Seit April 2019 bis heute (fortlaufend)
- **Zugänglichkeit:** Öffentlich und nutzbar für verschiedene Projekte
- **Regionale Fokussierung:** Unser Projekt konzentriert sich auf den Raum Berlin.

#### Datenbeschreibung:

Der SimRa-Datensatz enthält eine Vielzahl von Features, von denen folgende für unser Projekt besonders relevant sind:

Merkmal	Beschreibung
incidents	Anzahl der Beinaheunfälle in einem bestimmten Polygon (Straßensegment).
rides	Anzahl der Fahrten, die in einem bestimmten Polygon verzeichnet wurden.
score	Gefahrenbewertung basierend auf der Anzahl der Beinaheunfälle (Incidents) pro Segment im Verhältnis zu den Fahrten (Rides). Als besonders beängstigend markierte Vorfälle werden mit einem Faktor von 4,4 gewichtet.
geodaten	Die Geometrien sind als Polygone dargestellt, wobei jedes Polygon die oben genannten Merkmale enthält (siehe Abb.1)

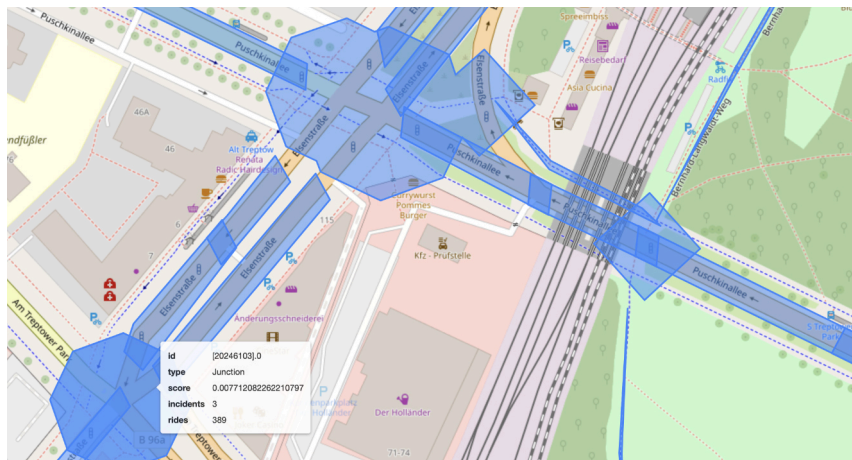


Abb.1: Visualisierung der Polygone aus dem SimRa-Datensatz (Screenshot aus dem Jupyter Notebook „simra\_data\_all.ipynb“)

## Mögliche Fehlerquellen:

Da es sich um ein Crowdsourcing-basiertes Projekt handelt, sind mögliche Fehlerquellen wie fehlerhafte oder unvollständige Nutzer\*innenangaben möglich. Auch technische Probleme der App wie fehlerhafte GPS-Daten oder Ungenauigkeiten bei der Erfassung von Vorfällen können auftreten.

## 2.2 OSM-Daten

Zusätzlich zu den SimRa-Daten wurden OpenStreetMap (OSM)-Daten verwendet, um wichtige Merkmale der Straßeninfrastruktur zu extrahieren. OSM bietet eine umfassende, global verfügbare, von der Community gepflegte und frei zugängliche Datenquelle.

Für unser Projekt konzentrieren wir uns auf das Fahrradnetzwerk (`network_type="cycling"`) von Berlin, um alle Straßenabschnitte zu identifizieren, die für Fahrräder nutzbar sind.

## Datenbeschreibung:

Es wurden folgende relevante Merkmale aus den OSM-Daten genutzt:

Merkmal	Beschreibung
highway	Klassifizierung der Straßenarten (z.B. Nebenstraße, Fußweg).
maxspeed	Maximale Geschwindigkeit, die auf Straßen zulässig ist
surface	Oberflächenbeschaffenheit der Straßen (z.B. Asphalt, Schotter).

## Mögliche Fehlerquellen:

- **Unvollständigkeit:** Daten können lückenhaft sein.
- **Qualität:** Die Genauigkeit der Daten variiert je nach Beitragenden.
- **Fehleranfälligkeit:** Mögliche Inkonsistenzen und Fehler durch verschiedene Nutzer\*innen.
- **Aktualität:** Verzögerungen bei der Aktualisierung realer Änderungen.
- **Subjektivität:** Unterschiedliche Beitragende könnten Merkmale wie Oberflächenbeschaffenheit unterschiedlich interpretieren.



## 2.3 Integration der Datensätze

Die Verknüpfung von SimRa- und OSM-Daten ermöglichte es, die Beziehungen zwischen Straßenmerkmalen und dem Gefahrenpotenzial zu untersuchen. Für die Modellierung wurden die Merkmale **highway**, **maxspeed** und **surface** als Prädiktoren verwendet, um potenzielle Einflussfaktoren auf den score (Gefahrenbewertung) zu identifizieren. Durch diese Integration sollten Muster und Korrelationen aufgedeckt werden, um das Gefahrenpotenzial auf Straßensegmenten präziser vorhersagen zu können.

## 3. Vorgehensweise

Unsere Vorgehensweise orientiert sich am *Cross-Industry Standard Process for Data-Mining (CRISP-DM)* und umfasst alle Phasen von der Problemdefinition über die Datenaufbereitung bis hin zur Modellierung und Evaluation. Entsprechend haben wir in sechs Sprints gearbeitet. (Die Jupyter Notebooks zu den einzelnen Sprints sind in der Datei projekt-verlauf.md verlinkt.)

### Sprint 1 - Verständnis für die Aufgabenstellung und Recherche von Datensätzen

Im ersten Schritt haben wir die Projektziele formuliert und intensiv nach geeigneten, kostenlosen Datensätzen gesucht. Da unser Vorhersagemodell auf Berlin fokussiert ist, interessierten uns nur lokal relevante Datensätze. Der **SimRa-Datensatz** wurde als Ausgangspunkt gewählt, ergänzt durch **OSM-Daten**, die eine kostenfreie, öffentlich zugängliche und relevante Grundlage für unser Projekt bieten. Außerdem stießen wir auf Datensätze zu Fahrradunfällen und zur Verkehrsdichte in Berlin, die für eine vertiefte Analyse ausgewählt wurden.

### Sprint 2 - Explorieren der Datensätze

Wir führten eine detaillierte Analyse der Datensätze durch, um deren Relevanz zu bewerten. Die Daten aus OSM wurden hierbei zunächst nicht explizit betrachtet, da uns noch eine geeignete Methode fehlte, um selektiv relevante Informationen aus OSM zu extrahieren. Bei der Analyse der Datensätze lag unser Fokus auf:

- **Datenmenge:** Ist sie ausreichend groß?
- **Features:** Welche spezifischen Merkmale sind enthalten?
- **Datenqualität:** Gibt es fehlerhafte oder unvollständige Daten?

Die Analyse ergab, dass die Daten zur Verkehrsdichte sowie zu den Fahrradunfällen für unsere Zwecke nicht nutzbar waren. Diese Datensätze enthielten entweder nur Daten für einen sehr begrenzten Zeitraum oder Messwerte für zu wenige geografische Punkte in Berlin, was ihre Aussagekraft erheblich einschränkte und sie somit als potenzielle Features ausschloss.

### Sprint 3 - Explorieren der OSM-Daten und Bereinigung unserer Datensätze

In diesem Sprint extrahierten wir relevante Daten aus den OSM-Daten. Ein erster Versuch über *overpass-turbo.eu* erwies sich als nicht erfolgreich. Über *Geofabrik* gelang es jedoch, das „*cycle\_network*“ herunterzuladen, welches ausschließlich Straßen umfasst, die für den Radverkehr geeignet sind, sowie sämtliche von OSM bereitgestellten Attribute für diese Straßen beinhaltet.



Die beiden Datensätze (SimRa und OSM) wurden umfassend analysiert und bereinigt:

### OSM-Daten

Besonders wichtig war es, alle 41 enthaltenen Features zu betrachten und zu bewerten, inwiefern sie die Sicherheit von Fahrradfahrenden beeinflussen könnten. Nach gründlicher Untersuchung entschieden wir uns, die folgenden drei Merkmale im OSM-Datensatz beizubehalten: **highway**, **maxspeed** und **surface**.

Der bereinigte Datensatz, der um nicht relevante Features reduziert wurde, wurde anschließend einer detaillierten Prüfung auf mögliche Fehler, fehlende Werte und die Verteilung der vorhandenen Attribute unterzogen.

- **maxspeed:** Die Analyse ergab, dass bei etwa 67 % der Straßen keine Höchstgeschwindigkeit hinterlegt war. Zudem gab es eine Mischung aus numerischen und nicht-numerischen Werten sowie unterschiedliche Einheiten (km/h und mph). Ähnliche Werte (z.B. 5, 6 und 7 km/h) wurden gruppiert, um eine konsistente Datenbasis zu schaffen. Fehlende Werte ergänzten wir basierend auf dem Straßentyp. Nach der Bereinigung wies der Datensatz keine fehlenden Werte mehr auf und enthielt sieben vereinheitlichte **maxspeed**-Kategorien.
- **surface:** Auch hier gab es fehlende Werte, die wir anhand des Straßentyps ergänzten. Durch das Zusammenfassen ähnlicher Oberflächentypen reduzierte sich die Anzahl der verschiedenen Oberflächenwerte von 36 auf 5 Kategorien. Doppelte Werte wurden entfernt, um Redundanzen zu vermeiden.
- **highway:** Es waren keine fehlenden Werte vorhanden, da der Straßentyp in OSM immer gesetzt ist. Dennoch reduzierten wir die Anzahl der Kategorien, indem wir selten vorkommende Straßentypen (weniger als 1 % Anteil) unter der neuen Kategorie „**highway\_rare**“ zusammenfassten.

Auffällig war der Straßentyp „**service**“, der Zugänge zu Parkplätzen, Toreinfahrten und ähnliche Wege umfasst. Hier waren ungewöhnlich wenige Werte gesetzt (siehe Abb.2), sodass wir hier eine gezielte Datenmanipulation durchführten.

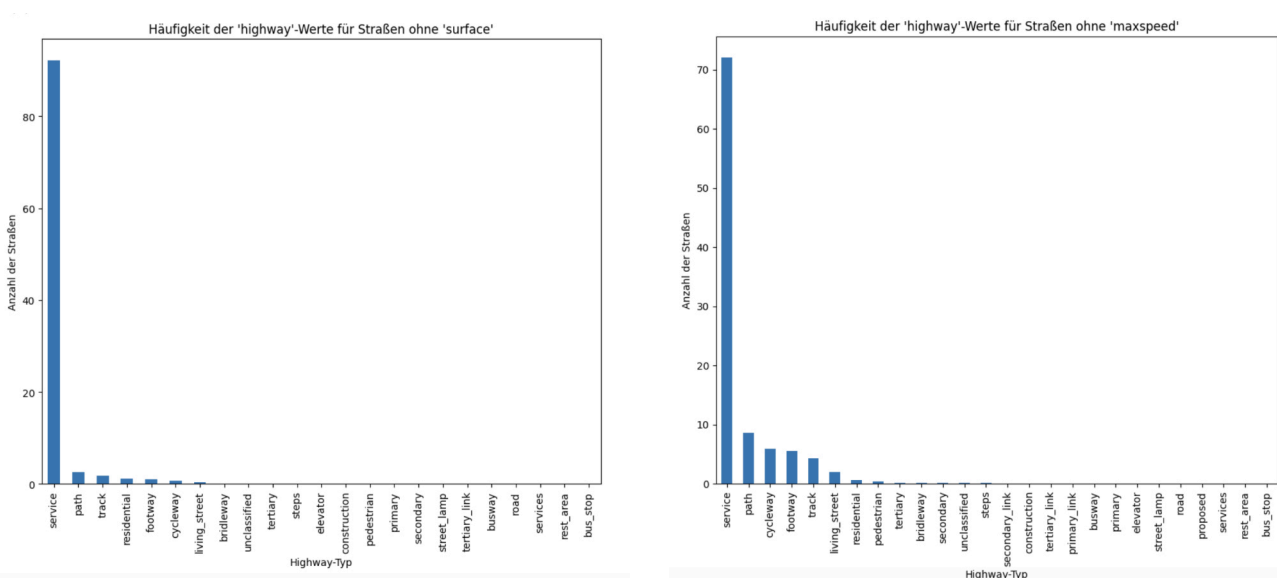


Abb.2: Verteilung fehlender maxspeed- und surface-Werte auf die einzelnen Straßentypen



Aufgrund dieser Auffälligkeit haben wir entschieden, mit zwei verschiedenen Versionen des OSM-Datensatzes weiterzuarbeiten: einmal mit den Service-Straßen enthalten und einmal ohne diese Straßen. So können wir beobachten, ob deren Einbeziehung einen Einfluss auf das Modellierungs- und Vorhersageergebnis hat.

Im Ergebnis stehen uns nun zwei bereinigte OSM-Datensätze zur Verfügung.

### SimRa-Daten

Die bloße Betrachtung der Anzahl der Beinaheunfälle reicht nicht aus, um aussagekräftige Schlüsse zu ziehen. Um die Daten sinnvoll zu interpretieren, ist es entscheidend, die Beinaheunfälle in Relation zu der Anzahl der dort aufgezeichneten Fahrten zu setzen (siehe Abb. 3). Dieses Verhältnis spiegelt sich im SimRa-Datensatz im Feature score wieder (siehe Abschnitt 2.1).

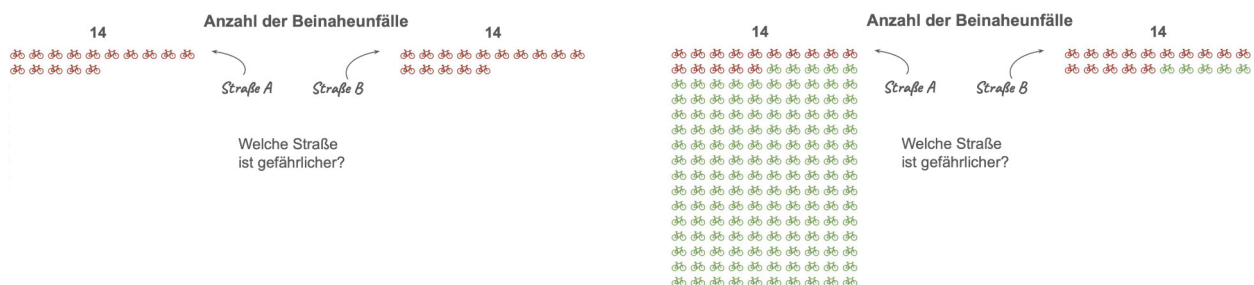


Abb.3 Relation der Beinaheunfälle zu der Gesamtzahl der gemachten Fahrten

Im Rahmen der Datenbereinigung untersuchten wir den Datensatz auf Duplikate, Ausreißer und fehlende Werte. Nach dem Entfernen von Duplikaten und Einträgen, die außerhalb Berlins erfasst wurden, verblieben von den ursprünglichen 16.819 Einträgen noch 16.294 valide Datensätze.

In diesem Sprint wurden die SimRa- und OSM-Datensätze so bereinigt und vorbereitet, dass sie im nächsten Schritt miteinander verknüpft werden konnten.

### Sprint 4 - Verknüpfung der Datensätze

Für das Modelltraining war es notwendig, die Datensätze SimRa und OSM zu verknüpfen. Dabei berücksichtigten wir mehrere Aspekte:

- Im SimRa-Datensatz sind die Geometrien als Polygone definiert, die mehrere Straßen umfassen und teilweise überlappen (siehe Abb. 1 bei 2.1 SimRa Datensatz). Dies führte zu Dopplungen der Werte im resultierenden Datensatz.
- Für das Feature **maxspeed** behielten wir den höchsten Wert und entfernten die niedrigeren, da wir annehmen, dass höhere Geschwindigkeiten ein größeres Gefahrenpotenzial für Radfahrende darstellen.
- Bei den Features **surface** und **highway** entschieden wir uns für eine prozentuale Verteilung der Werte auf das Polygon.
- Polygone im SimRa-Datensatz, die nicht mit OSM-Straßen verknüpft werden konnten, wie Rasenflächen, führten zu Zeilen ohne gesetzte OSM-Werte. Diese haben wir gelöscht.

Durch diese Bereinigung reduzierte sich der finale Datensatz auf etwa 15.500 Zeilen.

## Sprint 5 - Modelltraining

Wir haben vier verschiedene Modelle zur Evaluierung der Modellleistung getestet:

1. **k-Nearest Neighbors (k-NN)**
2. **Lineare Regression**
3. **Entscheidungsbaum (Decision Tree)**
4. **Random Forest**

Um die die Modelle zu optimieren, haben wir verschiedene Techniken angewendet:

1. **Hyperparameter-Tuning:** Hierbei haben wir **Grid Search** durchgeführt, um die optimalen Modellparameter für jedes Modell zu ermitteln. Diese Methode erlaubt eine systematische Suche nach den besten Parametern, um so die Modellperformance zu verbessern.
2. **Cross-Validation:** Um die Generalisierbarkeit unserer Modelle zu sichern, haben wir eine k-fache Cross-Validation eingesetzt. Diese Technik teilt den Datensatz in mehrere Teile und testet das Modell auf unterschiedlichen Datenpartitionen. Dadurch wird das Risiko der Überanpassung (Overfitting) reduziert.
3. **Feature Engineering:** Wir haben zusätzliche Merkmale extrahiert und irrelevante entfernt, um die Aussagekraft der Modelle zu erhöhen.
4. **Zweistufenmodell:** Diese Methode kombiniert verschiedene Klassifizierungsansätze, um die Vorhersagegenauigkeit zu steigern.
5. **Benutzerdefinierte Verlustfunktion (Custom Loss Function):** Diese Methode ermöglicht es, die Verlustberechnung gezielt an die Anforderungen der Anwendung anzupassen. Durch die Definition einer spezifischen Verlustfunktion können bestimmte Aspekte der Modellleistung optimiert werden, was potenziell zu besseren Trainingsresultaten führen könnte.

## Ergebnisse der Modelltests

Trotz dieser Maßnahmen konnten wir keine zufriedenstellenden Ergebnisse erzielen. Die folgenden Leistungswerte wurden gemessen:

	k-NN	Lineare Regression	Decision Tree	Random Forest
MSE	0.0011	0.00162	0.00256	0.0016
R <sup>2</sup> -Score	0.0430	0.00164	0.00089	0.0045

- **Mean Squared Error (MSE):** Dieser Wert gibt die durchschnittliche quadratische Abweichung zwischen den vorhergesagten und den tatsächlichen Werten an. Ein niedrigerer MSE-Wert weist auf eine bessere Modellleistung hin.
- **R<sup>2</sup>-Score:** Dieser Wert repräsentiert, wie gut die Streuung der Zielvariable durch das Modell erklärt wird. Ein R<sup>2</sup>-Score von 1 bedeutet, dass das Modell die Variabilität perfekt vorhersagt, während ein negativer Score darauf hinweist, dass das Modell schlechter als der Durchschnitt der Daten vorhersagt.



Der  $R^2$ -Score und der MSE sind wichtige Fehlermetriken zur Bewertung der Modellleistung. Ein bester  $R^2$ -Score von **0.0430** deutet darauf hin, dass das Modell kaum in der Lage ist, die Variabilität der Zielvariable zu erklären. Ein MSE von **0.0011** ist zwar relativ gering, dennoch muss dieser Wert im Kontext der Zielvariable betrachtet werden. Da die Zielvariable zwischen 0 und 1,5 liegt und der Großteil der Werte bei 0 konzentriert ist, könnte der MSE darauf hinweisen, dass das Modell vor allem die häufigsten Werte vorhersagt, während es die relevanten selteneren Werte ignoriert. Diese Werte zeigen, dass die Modelle die Daten unzureichend erklären konnten und somit nicht in der Lage waren, zuverlässige Vorhersagen zu treffen.

Hauptprobleme waren:

- Geringe Anzahl relevanter Daten
- Ungleichverteilung der Zielvariablen, der Gefahrenscores, (über 90 % der Werte bei 0), was es den Modellen erschwerte, Muster zu erkennen.

### Einfluss von Service-Straßen auf die Modellperformance

Die Analyse zeigte, dass die verschiedenen Modelle mit den unterschiedlichen Datensätzen – mit und ohne Service-Straßen – nur geringfügig unterschiedliche Ergebnisse lieferten. Dies deutet darauf hin, dass dieser Straßentyp keinen signifikanten Einfluss auf die Modellleistung hatte.

## Sprint 6 - Evaluation und Darstellung der Ergebnisse

Die Ergebnisse der Modellbewertung entsprechen nahezu unseren Erwartungen. Während der Datenaufbereitung sind wir bereits auf Herausforderungen gestoßen, die sich im Modelltraining bestätigt haben. Der finale Datensatz war relativ klein, was das Training erschwerte.

### Zentrale Probleme

**Unausgewogene Datenverteilung:** Die Zielvariable, der Gefahren-Score aus den SimRa-Daten, ist stark unausgewogen. Über 90 % der Werte liegen bei 0. Die verbleibenden Werte verteilen sich sehr nah bei 0, und der höchste erfasste Score liegt bei rund 1,5.

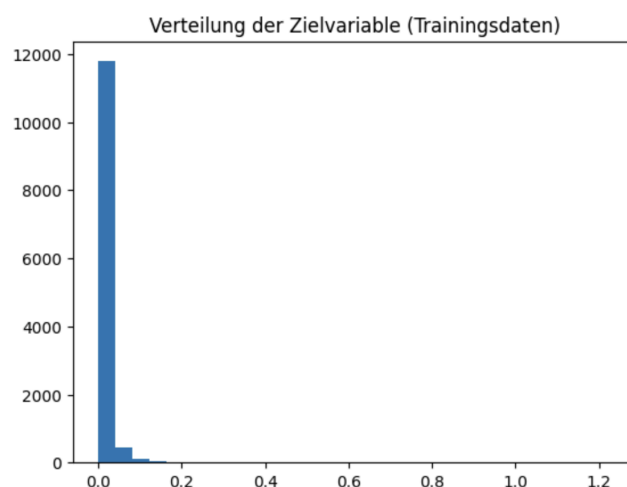


Abb.4 Verteilung der Zielvariablen

Die ungleiche Verteilung der Zielvariablen stellte eine erhebliche Herausforderung dar:

- **Erkennung von Mustern:** Aufgrund des stark unausgewogenen Datensatzes fällt es den Modellen schwer, nützliche Muster zu erkennen. Sie neigen dazu, die häufigsten Werte zu bevorzugen und somit die wenigen relevanten Datenpunkte zu ignorieren.
- **Modellleistung:** Die Werte der Fehlermetriken, wie  $R^2$ -Score und MSE, waren aufgrund dieser Ungleichheit niedrig, da die Modelle nicht in der Lage waren, die wenigen relevanten Fälle angemessen vorherzusagen.

## Ursachenanalyse für die schwache Performance

Um die Ursachen für die schwache Modellleistung besser zu verstehen, haben wir eine detaillierte Analyse durchgeführt. Die wichtigsten Gründe lassen sich wie folgt zusammenfassen:

1. **Repräsentativität der Daten:** Die geografischen und statistischen Daten waren möglicherweise nicht repräsentativ genug, um die Gefahrenstellen im Berliner Radverkehr präzise vorherzusagen.
2. **Feature Engineering:** Die ausgewählten Merkmale könnten unzureichend gewesen sein, um das Gefahrenpotenzial präzise zu prognostizieren.
3. **Modellkomplexität:** Das entwickelte Modell könnte zu einfach gewesen sein, um die Komplexität des Problems angemessen abzubilden.
4. **Underfitting:** Das Modell könnte zu simpel gewesen sein, um komplexe Muster zu erkennen.

Abschließend lässt sich sagen, dass die Herausforderungen hinsichtlich Datenverteilung und Modellkomplexität wertvolle Erkenntnisse für zukünftige Anpassungen und Verbesserungen bieten, um ein leistungsfähigeres Modell für die Vorhersage von Gefahren im Fahrradverkehr zu entwickeln.

## Visualisierung der Ergebnisse in der Streamlit-App

Da wir aufgrund der erwähnten Gegebenheiten kein funktionsfähiges Vorhersagemodell zur Verfügung hatten, konzentrierten wir uns auf die Visualisierung der Daten. Unsere Streamlit-App stellt die Features und die potentiellen Gefahrenpunkte auf einer Karte von Berlin dar. Dies ermöglichte uns, die extra-hierten OSM-Daten verständlich zu präsentieren.

## 4. Reflexion

### Erreichte Ziele

Trotz der Herausforderungen im Projekt haben wir wertvolle Erfahrungen im Umgang mit großen Datensätzen und Machine-Learning-Modellen gesammelt. Ein zentraler Erfolg war die sorgfältige Analyse und Datenbereinigung, die es uns ermöglichte, die Datensätze auf relevante Informationen zu reduzieren. Dies half uns, uns auf zwei wesentliche Datenquellen zu konzentrieren.

Obwohl die getesteten Modelle, wie k-NN, lineare Regression, Entscheidungsbäume und Random Forest, keine zufriedenstellenden Ergebnisse lieferten, war der Prozess lehrreich. Wir erhielten wichtige Einblicke in die Schwächen dieser Algorithmen im Kontext unserer Fragestellung und in die Herausforderungen bei der Vorhersage von Gefahrenpotenzialen für Radfahrende.

## Ausblick: Nächste Schritte

Ein zentraler Schritt zur Verbesserung des Projekts wäre die Integration weiterer externer Faktoren in die Analyse, wie beispielsweise Wetterdaten und aktuelle Baustelleninformationen. Diese zusätzlichen Daten könnten die Modellgenauigkeit erhöhen und umfassendere Vorhersagen ermöglichen. Die Anbindung an Echtzeitdaten würde nicht nur historische, sondern auch aktuelle Gefahrenlagen erfassen und somit gezieltere Vorhersagen liefern.

Für die Fortführung des Projekts könnten folgende Schritte sinnvoll sein:

- **Integration zusätzlicher Datenquellen:** Die Berücksichtigung von beispielsweise Wetterdaten oder Baustelleninformationen könnte die Modellgenauigkeit verbessern.
- **Erweiterung des Datensatzes:** Eine kontinuierliche Erhebung aktueller Daten und die Erweiterung des geografischen Fokus könnten die Aussagekraft erhöhen.
- **Feature Engineering ausweiten:** Zeitliche Daten, Wetterbedingungen, Informationen zu Kreuzungen und weitere infrastrukturelle Merkmale könnten die Modelle verbessern.
- **Fortgeschrittene Modelle testen:** Der Einsatz komplexerer Algorithmen oder sogar neuronaler Netze könnte helfen, nicht-lineare Zusammenhänge besser abzubilden.
- **Regularisierungstechniken anwenden:** Diese könnten helfen, ggf. Überanpassungen zu vermeiden und die Generalisierbarkeit der Modelle zu verbessern.
- **Modelloptimierung:** Weiteres Hyperparameter-Tuning könnte die Modellleistung zusätzlich verbessern.
- **Iterative Verbesserung:** Eine kontinuierliche Anpassung der Daten, Merkmale und Algorithmen könnte die Modellperformance schrittweise steigern.

Trotz der aktuellen Herausforderungen bietet das Projekt SafeCycling wertvolle Ansatzpunkte für zukünftige Verbesserungen. Durch weiterführende Datenintegration sowie Modelloptimierung könnten entscheidende Fortschritte in der Vorhersage von Gefahrenpotenzialen erzielt werden.