

```
1  # Getting started with Machine Learning (ML) and Support Vector Classifiers (SVC) - A
   systematic step-by-step approach
2
3  ## Abstract (de/en)
4  Wer sich mit dem Hypethema unserer Zeit ``Künstliche Intelligenz (KI)'' bzw. ``Machine
   Learning (ML)'' ernsthaft auseinandersetzen möchte, kommt nicht umhin, sich mit den
   grundlegenden ML-Algorithmen, entsprechenden Software-Werkzeugen, -Bibliotheken und
   Programmiersystemen zu beschäftigen.
5
6  Anyone who wants to seriously deal with the hypothetical topic of our time ``
   Artificial Intelligence (AI)'' or ``Machine Learning (ML)'' cannot avoid dealing with
   the basic ML algorithms, corresponding software tools, libraries and programming
   systems.
7
8  Allerdings wird jemand, der das erste Mal die Tür zu dieser gleichermaßen sehr
   spannenden wie beliebig komplexen und auf den ersten Blick unübersichtlichen Welt
   aufstößt, sehr schnell überfordert sein. Hier bietet es sich an, einführende und
   systematische Anleitungen zu konsultieren.
9
10 However, someone who opens the door for the first time to this equally very exciting
   as well as arbitrarily complex and, at first glance, confusing world will very quickly
   be overwhelmed. Here, it is a good idea to consult introductory and systematic
   tutorials.
11
12 Daher demonstriert das vorliegende Getting-Started-Tutorial anhand des sehr
   leistungsfähigen und performanten ``Support Vector Classifiers (SVC)'' sowie dem
   weithin bekannten und besonders anfängerfreundlichen ``Iris-Datensatz'' den typischen
   ML-Arbeitsprozess systematisch Schritt-für-Schritt.
13
14 Therefore, this Getting Started tutorial systematically demonstrates the typical ML
   work process step-by-step using the very powerful and performant ``Support Vector
   Classifier (SVC)'' and the widely known and exceptionally beginner-friendly ``Iris
   Dataset''.
15
16 Darüber hinaus werden die Auswahl des ``richtigen'' SVC-Kernels sowie deren Parameter
   beschrieben und ihre Auswirkungen auf das Klassifikationsergebnis werden gezeigt.
17
18 Furthermore, the selection of the ``correct'' SVC kernel and its parameters are
   described and their effects on the classification result are shown.
19
20 ## Introduction (de/en)
21
22 Von den **Arbeitsmitteln** in der **digitalisierten Arbeitswelt** wird immer stärker
   gefordert, dass sie sich selbstständig und aufgabenbezogen an sich ändernde
   Arbeitssituationen anpassen können. Diese **situative Adaptivität** kann je nach
   Stärke des Flexibilisierungsgrades oft nur durch Anwendung von **Artificial
   Intelligence (AI)** oder **Machine Learning (ML)** realisiert werden.
23
24 In the **digitised work environment**, there is an increasing demand for **Work
   equipment** to be able to adapt independently and in a task-related manner to changing
   work situations. This **situational adaptivity** can often only be realised through
   the use of **Artificial Intelligence (AI)** or **Machine Learning (ML)**, depending on
   the degree of flexibility.
25
26 Als Beispiele für solche KI-Anwendungen in der Arbeitswelt können vergleichsweise
   einfache **Sprachassistenzsysteme** (ähnlich z. B. Siri oder Alexa aus dem privaten
   Umfeld) bis hin zu teil- oder gar **vollautonomen Systemen** genannt werden. Solche
   vollautonomen Systeme sind beispielsweise sog. **fahrerlose Transportsysteme**, bei
   denen es sich um autonom fahrende Logistikfahrzeuge in größeren Industrieanlagen
   handelt.
27
28 Examples of such AI applications in work environments can range from comparatively
   simple **voice assistance systems** (similar, for example, to Siri or Alexa from the
   private sphere) to partially or even **fully autonomous systems**. Such fully
   autonomous systems are, for example, so-called **driverless transport systems**, which
   are autonomously driving logistics vehicles in larger industrial plants.
29
```

30 Neben den vielen sehr interessanten Vorteilen bzgl. Wirtschaftlichkeit, Arbeitserleichterung usw. kennzeichnet solche vollautonomen Systeme eine sehr hohe technische Komplexität. Diese betrifft sowohl ihre ****Betriebsfunktionen**** (z. B. autonome Navigation durch komplexe industrielle Umgebungen bei gemeinsamer Nutzung der Fahrwege durch andere menschlich gesteuerte Fahrzeuge) als auch seiner ****Sicherheitsfunktionen**** (z. B. Auswertung komplexer, miteinander verknüpfter, meist bildgebender Sicherheitssensorik zur Überwachung des Fahrraums).

31

32 In addition to the numerous very interesting advantages in terms of economic efficiency, workload reduction, etc., such fully autonomous systems are characterised by a very high level of technical complexity. This concerns both their ****operating functions**** (e.g. autonomous navigation through complex industrial environments with shared use of the roadways by other human-controlled vehicles) and their ****safety functions**** (e.g. evaluation of complex, interconnected, mostly imaging safety sensors for monitoring the driving space).

33

34 An solche autonomen Systeme und die hierfür eingesetzten KI-Algorithmen werden sehr hohe Anforderungen hinsichtlich der ****funktionalen Sicherheit**** gestellt. Jedoch sind die für eine sicherheitstechnische Bewertbarkeit notwendigen Anfordererungen hinsichtlich der ****Transparenz**** und ****Erklärbarkeit**** der durch KI getroffenen Entscheidungen je nach verwendeten KI-Algorithmen sehr schwer bis unmöglich. Beispielsweise werden durch aktuell laufende Forschungsprojekte die Transparenz und Erklärbarkeit von ****tiefen neuronalen Netzen**** untersucht. Weiterhin erfüllen heutige KI-Algorithmen hinsichtlich ihrer ****Erkennungsraten**** und damit ihrer ****Zuverlässigkeiten**** selbst unter günstigsten Bedingungen sehr oft nicht die Anfordererungen an die funktionale Sicherheit, um höhere Safety-Level (z. B. Performance Level d (PLd) nach ISO 13849) zu erreichen.

35

36 Very high requirements are placed on such autonomous systems and the AI algorithms used for this purpose with regard to ****functional safety****. However, depending on the AI algorithms applied, the requirements for the ****transparency**** and ****explainability**** of the decisions made by the AI are very difficult or even impossible to meet. For example, current research projects are investigating the transparency and explainability of ****deep neural networks****. Furthermore, today's AI algorithms, in terms of their ****recognition rates**** and thus their ****reliabilities****, very often do not meet the functional safety requirements to achieve higher safety levels (e.g. Performance Level d (PLd) according to ISO 13849), even under the most convenient conditions.

37

38 Eine hinsichtlich der geforderten funktionalen Sicherheit angemessene Bewertung oder gar ****Prüfung**** nach einheitlichen und idealerweise genormten Maßstäben hat viele Implikationen auf die zukünftige Ausrichtung des ****technischen Arbeitsschutzes**** in Deutschland und in Europa. Neben der derzeit noch sehr schwierigen algorithmischen Bewertbarkeit ist ein wichtiger Punkt, dass die bisherige klare Trennung zwischen ****Inverkehrbringensrecht**** (siehe z. B. Maschinenrichtlinie) und ****betrieblichem Arbeitsschutzrecht**** (siehe Arbeitsschutzrahmenrichtlinie und Betriebssicherheitsverordnung) so nicht mehr aufrechterhalten werden kann. Grund hierfür ist, dass sich die ****sicherheitsrelevanten Eigenschaften**** der autonomen Systeme durch während des Betriebs erlernte, neue oder ****angepasste Verhaltensweisen**** verändern werden.

39

40 An appropriate assessment or even ****testing**** with regard to the required functional safety according to uniform and ideally standardised criteria has many implications for the future orientation of technical ****occupational safety and health (OSH)**** in Germany and in Europe. In addition to the currently still very difficult algorithmic evaluability, an important point is that the previous clear separation between ****placing on the market law**** (see e.g. Machinery Directive) and ****occupational health and safety law**** (see European Occupational Health and Safety Framework Directive and German Ordinance on Occupational Safety and Health) can no longer be continued in this way. The reason for this is that the ****safety-relevant properties**** of the autonomous systems will change due to new or ****adapted behaviours**** learned during operation.

41

42 Aus diesen Gründen sollten sich insbesondere die zukünftig mit der Prüfung solcher Systeme befassten Akteure des technischen Arbeitsschutzes möglichst frühzeitig mit den KI- bzw. ML-Algorithmen vertieft auseinandersetzen. Nur dadurch lässt sich erreichen, dass die stürmische Entwicklung lernfähiger, adaptiver Systeme durch den

Arbeitsschutz und deren Prüfinstitute konstruktiv, kritisch und fachlich angemessen begleitet werden kann. Wird dies versäumt, muss aufgrund der Erfahrungen der vergangenen Jahre davon ausgegangen werden, dass das Arbeitsschutzsystem durch die wirtschaftlichen Interessen global agierender Softwaregiganten skrupellos umgangen oder ausgehebelt werden wird. Dies hätte die Folge, dass schwere oder tödliche Arbeitsunfälle wegen unzulänglich gestalteter KI-basierter Arbeitssysteme wahrscheinlich werden.

For these reasons, those involved in technical occupational safety and health who will be responsible for testing such systems in the future should familiarize themselves in depth with AI and ML algorithms as early as possible. This is the only way to ensure that the rapid development of adaptive systems capable of learning can be accompanied by OSH and their testing institutes in a constructive, critical and technically appropriate manner. If this is omitted, it must be assumed on the basis of the experience of recent years that the OSH system will be ruthlessly circumvented or undermined by the economic interests of globally operating software giants. This would have the consequence that serious or fatal occupational accidents are likely to occur due to inadequately designed AI-based work systems.

+++
@TODO: Bis hierhin gekommen.
+++

Wer einen ernsthaften fachlichen Einstieg in die Welt von ****Künstlicher Intelligenz (KI)**** bzw. ****Machine Learning (ML)**** sucht, wird nicht umhin kommen, sich mit den grundlegenden ML-Algorithmen, entsprechenden Software-Werkzeugen, Bibliotheken und Programmiersystemen auseinander zu setzen.

Anyone seeking a serious technical entrance into the world of ****Artificial Intelligence (AI)**** or ****Machine Learning (ML)**** will not be able to avoid dealing with the basic ML algorithms, corresponding software tools, libraries and programming systems.

Wer jedoch zum ersten Mal die Tür zu dieser ebenso spannenden wie beliebig komplexen und auf den ersten Blick verwirrenden Welt öffnet, wird sehr schnell überfordert sein. Hier empfiehlt es sich, einführende und systematische Anleitungen zu Rate zu ziehen.

However, someone who opens the door for the first time to this equally very exciting as well as arbitrarily complex and, at first glance, confusing world will very quickly be overwhelmed. Here, it is a good idea to consult introductory and systematic tutorials.

Ziel dieses Getting-Started-Tutorials ist es, den typischen ML-Arbeitsablauf systematisch und Schritt-für-Schritt am Beispiel des sehr leistungsfähigen ****Support Vector Classifier (SVC)**** zu demonstrieren.

The aim of this Getting Started tutorial is to systematically demonstrate the typical ML working process step-by-step based on the example of the very powerful and performant ****Support Vector Classifier (SVC)****.

Dieses Tutorial wird im Rahmen eines Workshops auf der DGUV-Fachtagung ****Künstliche Intelligenz**** voraussichtlich im November 2022 in Dresden vorgestellt. Der Workshop richtet sich an interessierte ML-Neulinge im technischen Arbeitsschutz der gesetzlichen Unfallversicherungsträger.

This tutorial will be presented as part of a workshop at the DGUV symposium ****Artificial Intelligence****, probably in November 2022 in Dresden. The workshop addresses interested ML novices in the technical occupational safety and health of the social accident insurance institutions.

Für die Zielgruppe des Workshops wurde der SVC-Algorithmus bewusst gewählt, um zu zeigen, dass es neben den ****tiefen neuronalen Netzen****, die in den Medien sehr präsent sind, noch viele andere sehr leistungsfähige ML-Algorithmen gibt. Andererseits wäre eine notwendige und verständliche Einführung in neuronale Netze und die technischen Hintergründe zu Perzeptoren, Aktivierungsfunktionen etc. für Neulinge in dem für den Workshop vorgegebenen Zeitrahmen nicht möglich gewesen.

68 For the target audience in the workshop, the SVC algorithm was intentionally chosen to
show that there are many other very powerful and performant ML algorithms apart from
the ****deep neural networks**** that are very present in the media. On the other hand, a
necessary and comprehensible introduction to neural networks and the the technical
background to perceptrons, activation functions etc. for newcomers would not be
possible within the time frame given for the workshop.

69

70 Außerdem befasst sich dieses Tutorial **nicht** mit der Erzeugung oder Akquisition von
ML-tauglichen Datensätzen. Der Grund dafür ist, dass ein ML-Neuling zunächst versuchen
wird (oder sollte), sich mit den ML-Algorithmen, Werkzeugen, Bibliotheken und
Programmiersystemen vertraut zu machen. Erst dann ist es sinnvoll, die eigene Umgebung
auf ML-taugliche Anwendungen hin zu untersuchen und daraus geeignete Datensätze zu
gewinnen.

71

72 Furthermore, this tutorial does **not** address the generation or acquisition of
ML-ready datasets. Reason for this is that a newcomer to ML will (or should) first try
to familiarize himself with ML algorithms, tools, libraries and programming systems.
Only then it makes sense to explore one's own environment with respect to ML-suitable
applications and to acquire suitable datasets from them.

73

74 Daher demonstriert dieses Tutorial die Verwendung ausgewählter ML-Tools in Form von
Python-Bibliotheken sowie die systematische Herangehensweise an den weithin bekannten
und sehr einsteigerfreundlichen ****Iris-Datensatz****. Laut Fachliteratur ist für die
Klassifikation des Iris-Datensatzes der Support Vector Classifier hinsichtlich
Erkennungsrate als auch Performanz besonders gut geeignet. Alternativ könnten auch
entscheidungsbaum-basierte ML-Algorithmen wie z. B. der
****Random-forests-Klassifikator**** eingesetzt werden.

75

76 Therefore, this tutorial demonstrates the usage of selected ML tools in the form of
Python libraries as well as the systematic approach to the widely known and very
beginner-friendly ****Iris dataset****. According to the literature, the Support Vector
Classifier is particularly well suited for the classification of the iris dataset in
terms of recognition rate and performance. Alternatively, decision tree-based ML
algorithms such as the ****Random Forests Classifier**** could be used.

77

78 Nach der Klassifikation des Iris-Datensatzes durch den SVC zunächst mit
Standard-Parametern wird darüber hinaus die Auswahl des "richtigen" SVC-Kernels mit
seinen Einstellparametern beschrieben und die Auswirkung auf das
Klassifikationsergebnis wird gezeigt.

79

80 After the classification of the iris dataset by the SVC initially with standard
parameters, the selection of the "correct" SVC kernel with its setting parameters is
furthermore described and the effect on the classification result is shown.

81

82