

Arbeit zur Erlangung des akademischen Grades
Bachelor of Science

**Development of a multivariate
algorithm for the classification of
B mesons at the LHCb experiment**

Nico Guth
geboren in Dinslaken

2022

Arbeitsgruppe Albrecht
Fakultät Physik
Technische Universität Dortmund

Erstgutachter: Prof. Dr. Johannes Albrecht
Zweitgutachter: PD Dr. Dominik Elsässer
Abgabedatum: 14.06.2022

Abstract

The precise measurements needed to further test the Standard Model of particle physics produce large amounts of data. To analyse this data, machine learning algorithms have gained popularity over the last two decades due to their high efficiency and generalizability. In particular, machine learning algorithms can help separate signal from background in studies where some types of background of neutral B_d or B_s mesons can be tough to identify using the signal decay kinematics alone. A classification algorithm is developed in this thesis for the distinction between neutral B_d and B_s mesons in pp -collisions. This algorithm uses data of tracks associated with the signal B meson without data of the signal decay. The machine learning models used in this thesis are trained on LHCb simulation. The chosen strategy is to first identify signal side fragmentation tracks using a Boosted Decision Tree classifier and then use the output of this classifier in a DeepSet classifier to identify the type of signal B meson. Evaluating this algorithm on a simulated sample shows a clear achieved separation between both B meson types. However, tests on LHCb data yield a substantially smaller separation.

Kurzfassung

Die präzisen Messungen, die nötig sind um das Standardmodell der Teilchenphysik weiter zu überprüfen, führen zu großen Datenmengen. Maschinelles Lernen hat über die letzten zwei Jahrzehnte immer mehr Beliebtheit in der Analyse solcher Daten erlangt. Im Speziellen können Algorithmen des maschinellen Lernens verwendet werden um Untergründe von neutralen B_d oder B_s Mesonen zu reduzieren. Das kann in manchen Studien hilfreich sein, wo die Trennung zwischen signal und Untergrund schwer mit der Kinematik des Signalzerfalls zu erreichen ist. Ein solcher Algorithmus wird in dieser Arbeit entwickelt um neutrale B_d von neutralen B_s Mesonen zu unterscheiden, welche in pp -Kollisionen produziert werden. Die Grundlage für diese Unterscheidung bilden Daten von Teilchenspuren, die in Verbindung mit dem Signal B Meson produziert werden, aber nicht aus dem eigentlichen Signalzerfall stammen. Die Daten zum Training der Algorithmen in dieser Arbeit sind simulierte LHCb Daten. Die gewählte Strategie ist es, als erstes Spuren der Signalseiten-Fragmentierung mithilfe eines Boosted Decision Tree Klassifizierers zu identifizieren, um dann die Ausgabe in einem DeepSet Klassifizierer zu benutzen der die Art des Signal B Mesons identifiziert. Die Evaluierung dieses Algorithmus auf simulierten Daten zeigt eine klare Trennung zwischen den beiden B -Meson Arten. Allerdings fällt die erreichte Trennung auf LHCb Daten deutlich geringer aus.

Contents

1	Introduction	1
2	Theoretical background	2
2.1	The Standard Model of particle physics	2
2.2	B meson production in proton-proton collisions	4
2.3	Classification algorithms	5
3	The LHCb detector	8
3.1	The LHC	8
3.2	The LHCb detector	8
4	Development of the B meson classifier	11
4.1	Choice of simulation samples for training	11
4.2	Same side track identification using a Boosted Decision Tree	12
4.3	Classification of B_d and B_s mesons	15
4.4	Testing the algorithm on LHCb data	19
5	Conclusion	24
	Bibliography	26

1 Introduction

Over the last centuries, experiments and techniques to probe physics on microscopic scales have become more and more precise. Nowadays, particle colliders such as the Large Hadron Collider (LHC) are used to probe the interactions of elementary particles. Studying the production and decay processes of subatomic particles in highly energetic particle collisions is a way to test our current understanding of particle physics. The Standard Model of particle physics has been proven to make accurate predictions on almost all observed subatomic processes. However, there are multiple physics phenomena this theory cannot explain. This includes the existence of dark matter and dark energy, the imbalance between matter and antimatter in the universe and even gravity. Therefore, further testing is still needed to find physics beyond the Standard Model.

The processes inside of particle collisions cannot be detected directly. Reconstructing the proton-proton collisions at the LHC is made possible by using particle detectors such as the LHCb detector. The LHCb detector is specialized on measurements of CP violation and rare decays involving b and c hadrons. Here, the products of a collision are analysed to identify the particles as well as the decays of these particles. Because of the complexity of this problem and the amount of data that needs to be analysed, in the last two decades more and more machine learning algorithms are used to support such measurements.

The goal of this thesis is to develop and train such an algorithm and evaluate its performance on LHCb data. The purpose of this algorithm is to distinguish between neutral B_d and B_s mesons in proton-proton collisions based on hadronization and fragmentation tracks while excluding tracks of the signal decay. The B_d mesons include B^0 and \bar{B}^0 mesons and the B_s mesons include B_s^0 and \bar{B}_s^0 mesons.

Such an algorithm could be used in other studies to reduce backgrounds from B mesons of another type than the signal. This also includes *partial backgrounds* where missing information of the background decay kinematics renders the identification of background almost impossible through simple cuts alone. Examples of such analysis modes are $B_s^0 \rightarrow K^+ \mu^- \nu$ where $B^+ \rightarrow J/\psi(\rightarrow \mu\mu)K^+$ is an unwanted and difficult to separate background or $B_s^0 \rightarrow D_s^+ K^-$ where B_d^0 and Λ_b^0 backgrounds leak into the signal region and complicate the measurement.

2 Theoretical background

The first section of this chapter introduces the basics of particle physics in a theory called the Standard Model. The next section describes the B mesons relevant to this thesis and the processes in a proton-proton collision. The last section briefly explains the machine learning classification algorithms used in this thesis.

2.1 The Standard Model of particle physics

Physics phenomena on the smallest observed length scales are described by a quantum field theory known as the Standard Model (SM). The SM contains all known elementary particles and the interactions between them. The elementary particles are separated into twelve fermions with spin $\frac{1}{2}$, four gauge bosons with spin 1 and the Higgs boson with spin 0. Each of those particles has a corresponding antiparticle with all of its charge-like properties inverted. Neutral particles can be identical to their antiparticles. While fermions make up all the currently observable matter in the universe, bosons describe the interactions and properties of every particle. The following paragraphs briefly introduce each fundamental particle listed in table 2.1.

Table 2.1: Listing of all fundamental particles in the Standard Model.

fermion						boson				
quark			lepton			gauge-boson				scalar-boson
u	c	t	e^\pm	μ^\pm	τ^\pm	γ	g	Z	W^\pm	H^0
d	s	b	ν_e	ν_μ	ν_τ					

The atomic matter in the universe consists of mainly three of the elementary fermions. These are the up quark with charge $\frac{2}{3}e$, the down quark with charge $-\frac{1}{3}e$ and the electron with charge $-1e$. The up and down quarks are the valence quarks of the protons and neutrons inside atomic nuclei. Additionally, processes of the weak interaction often involve the neutrally charged electron-neutrino. These four fermions, together with their antiparticles, make up the so-called first particle generation. The second and third particle generations have the same structure as the first generation, but the masses of the corresponding particles increase with each generation. Each SM generation has one up-type quark (up u , charm c , top t), one down-type quark (down d , strange s , bottom b), one charged lepton (electron e , muon μ , tau τ) and one chargeless lepton (electron-neutrino ν_e , muon-neutrino ν_μ ,

tau-neutrino ν_τ). Originally, neutrinos were assumed to be massless in the SM, but this changed with the discovery of neutrino oscillations [1, 2].

The interactions between all particles are described by the four fundamental forces (electromagnetism, the strong force, the weak force, gravity). While the SM does not include gravity, the gauge bosons are the mediators of the three other forces.

The photon γ is charge- and massless, and it mediates the electromagnetic force. The electromagnetic force is described in quantum electrodynamics (QED) and couples to the charge Q of a particle.

The gluon g is also charge- and massless, and it mediates the strong force. The strong force is described in quantum chromodynamics (QCD) and couples to the color-charge (red, green, blue, anti-red, anti-green, anti-blue) of a particle. Only quarks and gluons are color-charged. This means, in contrast to QED, the force-carrier particles in QCD can interact with themselves. Also, the force between two color-charged particles increases with their distance. This leads to the so-called color-confinement, which states that only neutrally color-charged particles can exist in isolation. A particle is color neutral if either a color is canceled with the corresponding anti-color or all three colors are present in equal amounts. This prohibits the direct observation of quarks, and only collections of quarks, called hadrons, are observable. There are multiple classes of hadrons, but the most common classes are mesons (one quark and one antiquark) and baryons (three quarks). Tetraquarks and pentaquarks were first observed at LHCb in 2015 and 2016 [3, 4]. The bound state of neutrons (udd) and protons (uud) in the atomic nucleus is also the result of the strong force.

The mediators of the weak force are the chargeless Z -boson and the W^\pm -boson with charge $\pm 1e$. The W^\pm boson couples to the third component of the weak isospin $T_3 = \pm \frac{1}{2}$. The Z boson couples to the weak hypercharge $Y_W = 2(Q - T_3)$. Charged currents through the W^\pm boson are the only known way to change the flavour of a particle. There are six quark flavours and six lepton flavours matching the types of fermions listed in table 2.1. This results in processes such as the β^- -decay ($n \rightarrow p e^- \bar{\nu}_e$) or the β^+ -decay ($p \rightarrow n e^+ \nu_e$).

The Higgs boson H^0 is chargeless, and it is an excitation of the Higgs field. Through the interaction with the Higgs field all fermions as well as the W^\pm and Z bosons gain their masses. The Higgs boson is the last experimentally found elementary particle and was first observed at ATLAS and at CMS in 2012 [5, 6].

2.2 B meson production in proton-proton collisions

A proton-proton collision (pp -collision) is the interaction between two highly energetic protons through the strong force. The exchange of gluons between those protons initiates the production of multiple particles. Important at LHCb are primarily pp -collisions in which b and c hadrons are produced. The dominant process for b hadron production is the fusion of two gluons into one gluon, which then decays into a $b\bar{b}$ pair. These b quarks then hadronize by combining with other quark pairs from the vacuum. Additionally, fragmentation particles are produced through further hadronization of the unpaired quarks.

A frequent goal at LHCb is to analyse pp -collisions for B meson candidates. B mesons are composed of one b quark and one of the $u/d/c/s$ quarks. Relevant for this thesis are the neutral B mesons

$$B^0 = \bar{b}d, \quad \bar{B}^0 = b\bar{d}, \quad B_s^0 = \bar{b}s, \quad \bar{B}_s^0 = b\bar{s}.$$

The algorithm developed in this thesis distinguishes B_d mesons (B^0 and \bar{B}^0) from B_s mesons (B_s^0 and \bar{B}_s^0). An example schematic of a pp -collision containing a B meson is shown in fig. 2.1. This schematic excludes particles which are unassociated with the signal B that are contained in the underlying event.

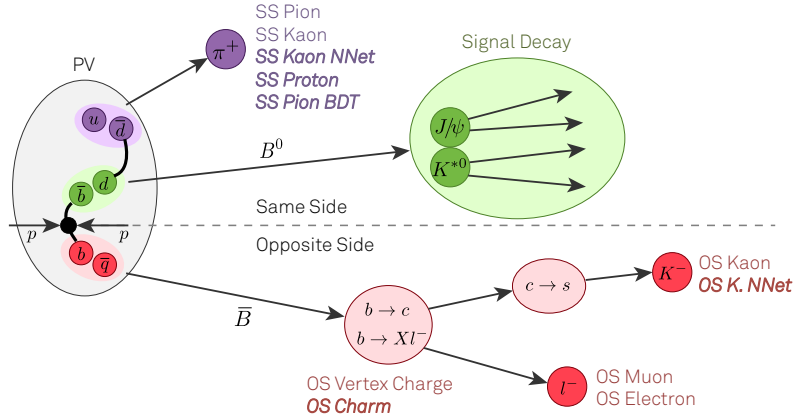


Figure 2.1: Schematic overview of the underlying principles of LHCb's flavour tagging algorithms [7].

When a B meson in a pp -collision is Lorentz boosted, it is possible to distinguish its production point from its decay point. This allows matching of the decay particles to the signal B . The production point, i.e. the collision point of the pp pair, is called the *primary vertex* (PV) and the decay point is called the *secondary vertex* (SV). Identifying the type of signal B is possible by reconstructing either the signal decay

or the associated event. The latter method is independent of the signal decay channel, and it is often used in flavour tagging algorithms at LHCb. Flavour tagging is a technique for distinguishing matter from antimatter states of B mesons using the associated event. The method developed in this thesis uses similar concepts but aims at B_d from B_s mesons instead.

The back-to-back production of the $b\bar{b}$ pair leads to an angular separation of the associated event into two sides. The *same side* (SS) contains the decay particles and the fragmentation particles of the signal B . The *opposite side* (OS) contains all particles associated with the b quark that is not in the signal B . For the distinction between B_d and B_s mesons, it is irrelevant if the signal B contains a b quark or an \bar{b} quark. Instead, the light quark partner of the b quarks needs to be identified. Therefore, it is expected that only the SS fragmentation contains relevant information for this analysis.

2.3 Classification algorithms

The objective of any classification algorithm is to assign a class to each given data point so that some property is shared among all data points in a class. For example, the main algorithm of this thesis evaluates the measured data of a pp -collision to estimate the type of the signal B meson. All classification algorithms used in this thesis are *supervised machine learning algorithms*. Such algorithms optimize the parameters θ of a given model f so that the class predictions $\hat{y} = f(x|\theta)$ best fit to a given target y . The process of optimizing these parameters is called *training* and often involves minimizing a given *loss function* $L(\hat{y}|y)$ that describes how the prediction \hat{y} compares with the target y . The following sections describe the classification algorithms that are used in this thesis.

2.3.1 Boosted Decision Trees

A *Boosted Decision Tree* (BDT) is an ensemble of multiple different *decision trees* trained and evaluated on the principles of *gradient boosting*. Decision trees are *binary trees* with a decision condition at each node and a prediction score for each leaf. The class prediction made on a given data point is based on a series of binary decisions on single features starting at the root and ending in a leaf of the decision tree. The weighted sum of the prediction scores of all trees in a BDT is then transformed using the *logistic function* to resemble an estimated probability that the given sample belongs to one of the classes. In gradient boosting the decision trees are trained iteratively, so that at each step the weighted sum of all previous decision trees and

the current decision tree minimizes a given loss function. The gradient of the loss function with respect to the model parameters is used to find this minimum.

2.3.2 Neural Networks

A *Neural Network* (NN) is a non-linear transformation of an input vector \vec{x} into an output vector \hat{y} . This transformation is divided into multiple steps called layers. The n th layer represents a vector $\vec{a}^{(n)}$, also called the *activation*, that is calculated based on the values of the previous layer using the formula

$$\vec{a}^{(n)} = f^{(n)} \left(W^{(n)} \cdot \vec{a}^{(n-1)} + \vec{b}^{(n)} \right). \quad (2.1)$$

The *weights*-matrix $W^{(n)}$ and the *bias*-vector $\vec{b}^{(n)}$ are parameters of the model that have to be adjusted in the training process. The *activation function* f can be any differentiable function that should be non-linear. Used in this thesis are the activation functions

$$f_{\text{ReLU}}(z) = \max(0, z) \quad \text{and} \quad (2.2)$$

$$f_{\text{Sigmoid}}(z) = \frac{1}{1 + e^{-z}}. \quad (2.3)$$

The first layer of a NN is the input layer with the activation $a^{(0)} = \vec{x}$ and the last layer is the output layer with the activation used as \hat{y} . All other layers are called *hidden layers*.

The iterative training of a NN uses a technique called *backpropagation*. As a first step, all weights of the NN are set to random values. In each iteration of the training, the output of the NN is evaluated to estimate the gradient of the loss function with respect to all weights using the chain rule. Based on this gradient, the weights are adjusted in the direction of the minimum of the loss function with a step size that is set by an *optimizer*. This thesis uses the loss function *binary cross entropy* in combination with the optimizer *Adam*[8].

To reduce the *overtraining* of the NN and increase its generalizability, this thesis uses regularization techniques called *early stopping* and *dropout*. In early stopping, the performance of the model is evaluated on a validation dataset after each iteration, and the training is stopped when this performance does not increase over N iterations. This way, further adjusting of the NN to the training data is prevented, and only the best model with respect to the validation data is used. In dropout, at each training iteration a smaller subset of the NN is trained by temporarily setting random rows in W and \vec{b} to 0. The number of rows that are being disabled has to be set for each layer before training the NN. Dropout leads to a more robust NN by reducing the co-dependence of pairs of activation elements.

2.3.3 DeepSets

A *DeepSet*[9] is an extension of NNs to allow inputs of sets of vectors. DeepSets can therefore be used to solve problems where the data of each sample contains a variable length list and where the solution should be invariant under permutations of this list. In this thesis, a DeepSet is used to classify *pp*-collision events that contain different amounts of tracks of which the order in the data should not matter.

The underlying concept of a DeepSet is the function

$$f(X) = \rho \left(\sum_{x \in X} \phi(x) \right) \quad (2.4)$$

which is invariant under permutations of the instances x in a set X . ρ and ϕ represent NNs and can be vector-valued. In principle, the sum could be replaced with any permutation invariant function, but in this thesis only sums are used. In other words, a DeepSet contains a ϕ -network and a ρ -network. The ϕ -network first extracts some features about each instance in the set. The extracted feature values of all instances are then summed up and propagated into a ρ -network that calculates the final output. Like NNs, DeepSets are also trained through backpropagation and the process is fully analogous to the training of a NN.

3 The LHCb detector

The data used in this thesis originates from the LHCb detector, which is located at the LHC. Therefore, this chapter first introduces the LHC and then describes the LHCb detector.

3.1 The LHC

The LHC (Large Hadron Collider) is a 27 km long ring accelerator located in the vicinity of Geneva at the border of France and Switzerland. It is the largest particle accelerator in the world, and it is managed by CERN (European Organization for Nuclear Research). The LHC can collide protons with a center-of-mass energy of up to 14 TeV at a peak luminosity (interactions per area and time) of $10^{34} \text{ cm}^2 \text{ s}^{-1}$ [10]. Additionally, the LHC can also accelerate heavy ions. The four largest detectors at the LHC are ATLAS, CMS, ALICE and LHCb. While ATLAS and CMS are general-purpose particle detectors, ALICE is dedicated to heavy ion collisions resulting in a quark-gluon plasma.

3.2 The LHCb detector

The LHCb detector is a particle detector dedicated to pp collisions involving c or b quarks. The goals of the LHCb collaboration are precision measurements of CP -violation and rare decays. To accomplish these goals, the LHCb detector has been built as a single arm forward-spectrometer covering angles to the beam pipe from 10 mrad to 300 mrad [11]. This design has been chosen because the majority of high energetic b -/ c -hadron pairs, produced in pp collisions, have Lorentz boosted velocities in the same direction as one of the incident protons. An overview of the detector is shown in fig. 3.1. The following paragraphs briefly describe the components of the LHCb detector based on the article [11].

A particle produced at the PV first traverses the *Vertex Locator* (VELO). There, the tracks of charged particles are measured using silicon semiconductor sensors at a resolution of about 4 μm . This information is primarily used to reconstruct the position of the primary and secondary vertices of the decay particles.

Beyond the VELO, the *Tracking System* then further measures the track points of the same particles. This is done in the *Tracker Turicensis* (TT) before the magnet and in the Trackers T1, T2 and T3 behind the magnet. The dipole magnet

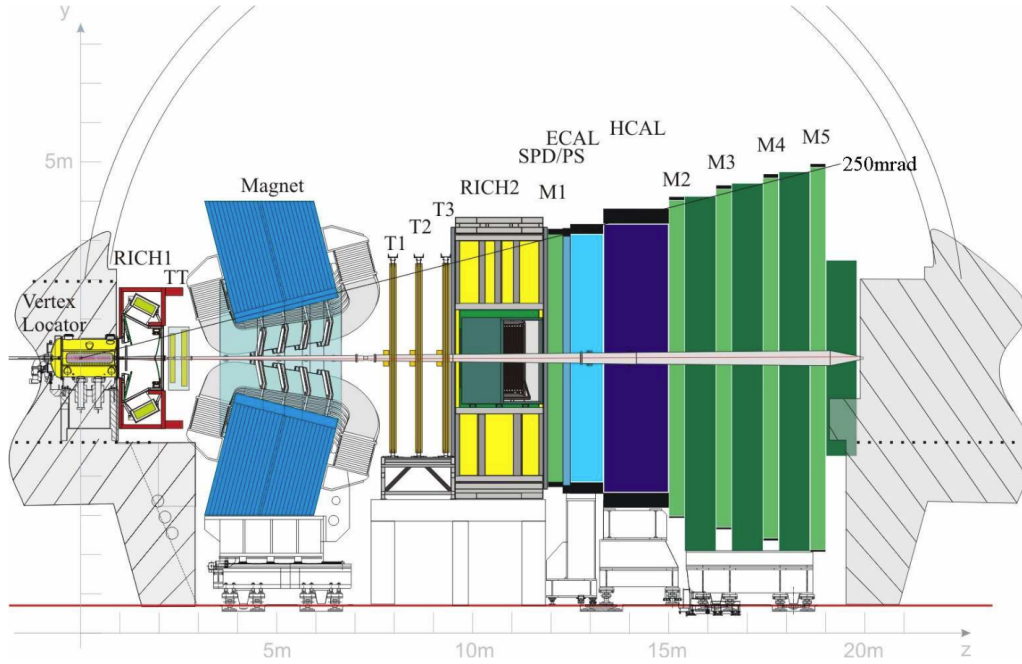


Figure 3.1: Sideview of the LHCb detector [11].

bends the particle tracks through the Lorentz force so that the particle's momentum can be inferred. While the TT and the inner parts of the T1-T3 also use silicon semiconductor sensors, the outer parts of the T1-T3 use straw-tube modules which are gaseous ionization detectors.

The *Ring Imaging Cherenkov detectors* (RICH1, RICH2) are used to measure the particle velocities. When charged particles traverse optically dense materials, the emission of Cherenkov radiation can be induced. This radiation is then detected by photomultipliers. The opening angle of the emitted light cone is dependent on the velocity of the particle. The use of multiple radiators allows measurement of particles with momenta between 1 GeV and 100 GeV. Comparing the momentum measured in the Tracking System with the velocity measured in the RICH detectors yields information particle's mass.

The *calorimeters* (ECAL and HCAL) absorb and measure the energy of most particles. In dense materials, particles can induce particle showers with the size directly related to the energy deposited. Each calorimeter is structured in alternating layers of absorbers and scintillators. The *electronic calorimeter* (ECAL) absorbs electrons, positrons and photons using lead, and the *hadronic calorimeter* (HCAL) absorbs any hadrons using iron. The calorimeters are the only detectors at LHCb

which can detect chargeless particles.

The *muon chambers* (M1-M5) track charged particles which pass through the calorimeters without absorption. The majority of charged particles at this point in the detector are muons. The iron layers between the muon chambers are used to absorb any non-muon particles. M1-M5 are *multi-wire proportional chambers* to detect the track points of the muons.

The combined measured data of the RICH, the calorimeters and the muon chambers is used for particle identification. Together with the Tracking System, this allows the reconstruction of the production and decay processes related to the *pp*-collision.

Due to the high *pp*-collision rates of the LHC, only a small portion of the *pp*-collisions measured by the LHCb detector is permanently recorded for further analysis if it passes the LHCb trigger requirements. Therefore, a Trigger System is implemented to reduce the event detection rate from about 10 MHz to about 2 kHz. This system consists of the hardware trigger *Level-0* (L0) and the software trigger *High Level Trigger* (HLT). The L0 collects information from the pile-up system of the VELO, the calorimeters and the muon chambers. If the event does not contain particles with high transversal energy or momentum, it is rejected. The L0 allows the readout of the whole detector at a rate of about 1 MHz. The HLT then uses the data of all detector components to fully reconstruct the event and select events on various decision criteria. At the final rate of about 2 kHz, the event data is permanently recorded and can then be accessed on demand.

4 Development of the B meson classifier

The goal of this thesis is to develop an algorithm that can distinguish B_d mesons from B_s mesons produced in pp -collisions. This algorithm is intended to be independent of the signal decay channel. Therefore, it only uses data of the associated event excluding tracks from the signal decay. As mentioned in section 2.2, only the SS is expected to contain relevant information. Thus, every track in the dataset is assigned an estimated probability of being a SS track by a BDT before classifying the signal B using a DeepSet.

This chapter describes in detail the methods and results of training this B meson classifier. The first section explains the choice of the training dataset. Then the training of a BDT for SS track classification is presented. The following section describes the training of a DeepSet for B meson classification. In the last section, the trained model is tested on LHCb data.

4.1 Choice of simulation samples for training

The training of supervised classification algorithms requires data in which the true classes are known. For the analysis described in the following, classification algorithms are exclusively trained with LHCb simulation.

Since both the B_d and B_s mesons are needed for the training of a B_d and B_s classifier, data of two different signal decay channels can be combined in this thesis ($B_d \rightarrow J/\psi K^*$ and $B_s \rightarrow D_s^+ \pi^-$). Multiple datasets are available for both signal decays. It is important to choose datasets that are compatible to each other so that the trained models do not learn artificial differences, that are not found in real detector data. Therefore, the available datasets are compared by plotting histograms of each feature. On the basis of this comparison, multiple sources of differences between the datasets are found. The differences are eliminated by choosing only datasets that originate from simulations of the same year (2016) generated with the same simulation software version. Furthermore, identical selection requirements are used for particles both decay channels have in common.

4.2 Same side track identification using a Boosted Decision Tree

To identify tracks that belong to the SS of a signal B , a BDT is trained on the simulated LHCb data. The BDT output resembles an estimated probability of the track being in the SS and is from now on called Prob_{SS} . The implementation of the BDT is done with the library XGBOOST [12].

The dataset used for training the BDT contains about 70 different features of the tracks of the associated event. To reduce this list of features, first the correlation coefficient between all feature pairs is calculated. Features with a correlation near 100% are considered to contain redundant information. Only one feature of each set of highly correlated features is kept for training the BDT. Next, a BDT is trained on a dataset containing all features and the gain of each feature is calculated. The gain is a measure used in decision trees to estimate the accuracy improvement of the introduction of a cut on one feature. Additionally, the permutation importance of the ROC AUC is calculated. The permutation importance is the difference in any performance metric after making predictions on a dataset with the entries of a single feature permuted at random. The features with the lowest gain or permutation importance are discarded until a reasonable amount of features is achieved. The remaining 21 features are listed in table 4.1 together with the corresponding feature importances shown in fig. 4.1.

Table 4.1: List of all features used to train the BDT for SS track identification.

feature	feature
p_T	IP_{SV}
p_{proj}	$\chi^2(\text{IP}_{\text{SV}})$
Δp_T	$\sigma(\text{IP}_{\text{pileup vtx}})$
Δz	$\text{IP}_{\text{best PV}}$
$\Delta \eta$	$\chi^2(\text{IP}_{\text{best PV}})$
$\cos(\Delta \phi)$	IP_{min}
$\text{Prob}_{\text{ghost}}$	same PV
$\chi^2(\text{vtx})$	cone isolation
SumBDT	$N_{\text{non iso}}$
MinBDT	$\sum p_{\text{in cone}}$
SumMinBDT	

All the following features describe a single particle or rather its reconstructed track. The transversal momentum is denoted by p_T . Some features are also defined for

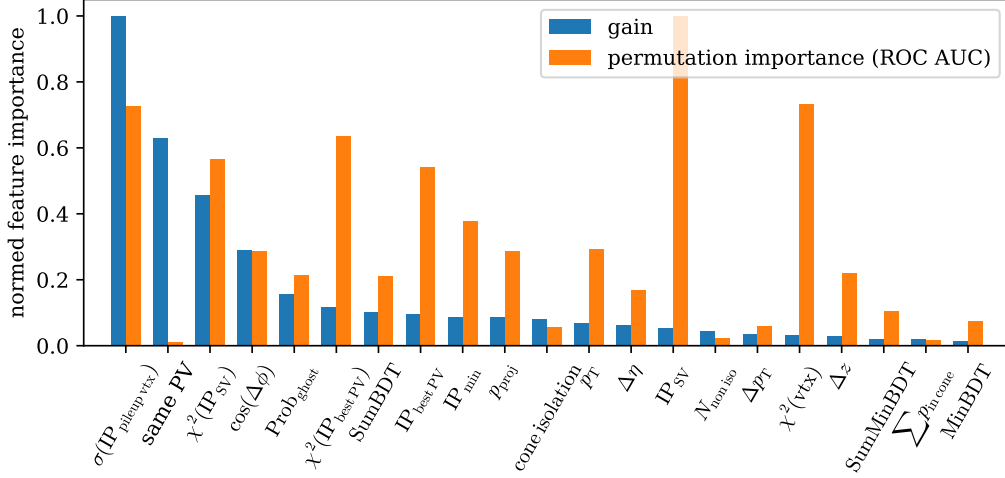


Figure 4.1: Calculated feature importances on the trained BDT for SS track identification. Shown are the gains and the permutation importances of the ROC AUC normed on the largest value.

relation between the signal B and the reconstructed track. p_{proj} is the dot product of the 4-momenta, Δp_T is the difference of the transversal momenta, $\Delta\eta$ is the difference of the pseudorapidities and $\cos(\Delta\phi)$ is the cosine of the difference of the ϕ -coordinates. Δz is the z -coordinate difference of the PV and the first detector hit of the track. $\text{Prob}_{\text{ghost}}$ is an estimated probability that the track is falsely reconstructed and does not belong to an actual particle. $\chi^2(\text{vtx})$ is the χ^2 of the combined reconstructed vertex of the signal B and the track. “same PV” refers to whether the particle’s vertex is the same as the PV. An impact parameter (IP) is the minimum distance of a track to a given point. IP_{SV} is the IP to the SV and $\chi^2(\text{IP}_{\text{SV}})$ is its uncertainty. $\sigma(\text{IP}_{\text{pileup vtx}})$ is the significance of the IP to the closest pileup vertex, $\text{IP}_{\text{best PV}}$ is the IP to the best estimated PV and $\chi^2(\text{IP}_{\text{best PV}})$ is its uncertainty. IP_{min} is the minimum IP with respect to all PV candidates. Cone-isolation refers to an isolation metric of a track which is evaluated based on how many neighbouring tracks are inside a defined cone in the η - ϕ -plane around the track in question. SumBDT, MinBDT and SumMinBDT describe the cone-isolation based on multiple BDT classifiers which work on the mentioned principle. $\sum p_{\text{in cone}}$ is the sum of the absolute momenta of all particle tracks in the defined cone. The feature “cone isolation” describes the momentum based cone-isolation $p/\sum p_{\text{in cone}}$. $N_{\text{non iso}}$ is the number of tracks not isolated from the track.

Then a BDT is trained on a dataset containing all features of this list. From the 18 million tracks in the dataset, 60% are used for training and 40% are used

for validation of the trained model. The trained model contains 2000 estimators at a maximum decision tree depth of 4. The learning rate is set to 0.1. Due to the imbalance of SS tracks and other tracks in the data ($N_{\text{SS}}/N_{\text{other}} \approx 0.084$), a parameter that controls the balance of positive and negative weights of the BDT is set accordingly. The learning objective is logistic regression for binary classification.

The negative log-likelihood and the error rate for each iteration of the training of the BDT are shown in fig. 4.2. The error rate is the proportion of predictions matching the true labels of the simulation. For this, all tracks with $\text{Prob}_{\text{SS}} > 0.5$ count as predicted SS tracks.

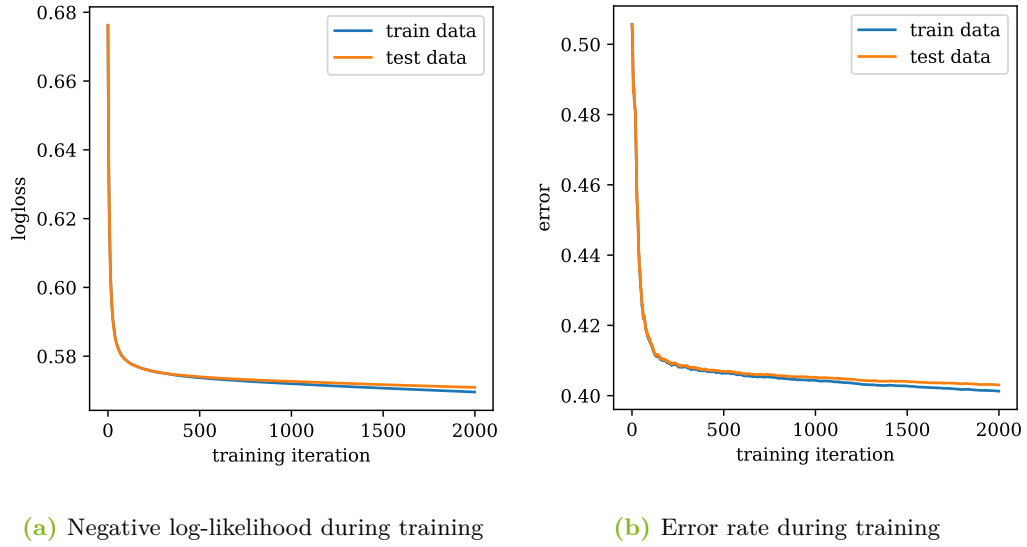


Figure 4.2: Performance of the BDT for SS track identification after each training iteration.

To show the achieved separation of the SS tracks, histograms of Prob_{SS} split by the true labels of the simulation are shown in fig. 4.3a. A measure of separation is the ROC (receiver operating characteristic) curve shown in fig. 4.3b. To estimate the generalization and overtraining of the model, each performance measure is calculated on both the test data and training data. The achieved ROC AUC (area under the ROC curve) is 0.763 on the test data and 0.767 on the training data. A ROC AUC of 1.0 means perfect separation and a ROC AUC of 0.5 means no separation.

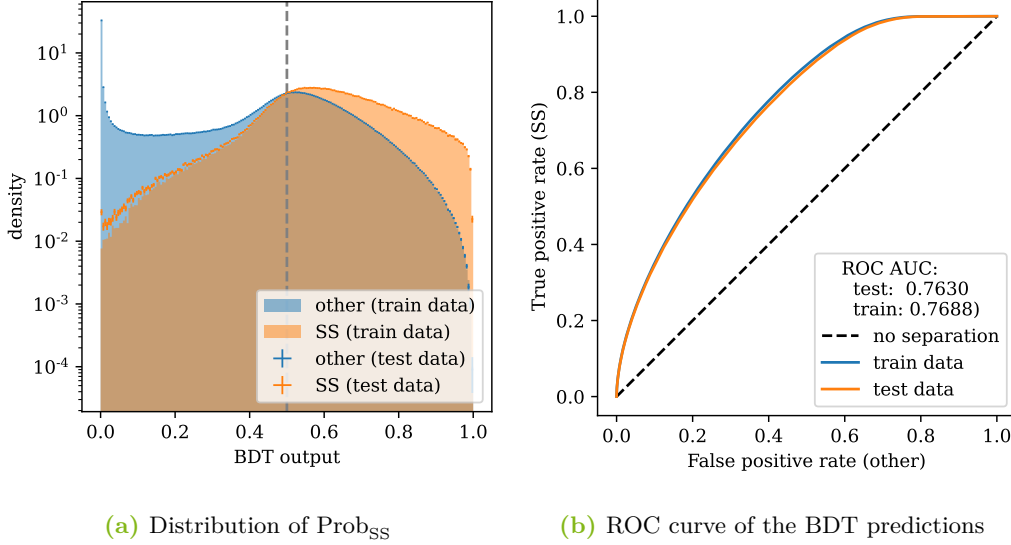


Figure 4.3: The left figure shows the distribution of the BDT output split by the true labels of the simulation. The right figure shows the ROC curve of the BDT output. Both figures show the BDT prediction for the test data and the training data.

4.3 Classification of B_d and B_s mesons

The distinction between B_d and B_s mesons based on the associated event is a non-trivial problem. Since both mesons have different masses and quark contents, there are difference in the kinematics and types of particles in the associated event. Therefore, in this thesis it is tested whether a multivariate machine learning algorithm can spot these differences. The algorithm of choice for this thesis is a DeepSet that is trained on simulated LHCb data. As explained in section 2.3.3, a DeepSet can handle data of variable length where the order of its elements should not change the result of the classification. Therefore, DeepSets are ideal to classify pp -collision events that contain different amounts of particle tracks. The DeepSet trained here, outputs an estimated probability Prob_{B_s} for each event that the signal B is a B_s meson. Therefore, $1 - \text{Prob}_{B_s}$ is the estimated probability that the signal B is a B_d meson. The implementation of the DeepSet is done with the library PYTORCH [13].

The procedure of selecting important features is similar to the procedure described in section 4.2. However, a DeepSet was trained on all features and the feature importances used here are the permutation importance of the ROC AUC and the

permutation importance of the accuracy. Again, features with the lowest feature importance scores are discarded until a reasonable amount of features is achieved. The remaining 23 features are listed in table 4.2 and the corresponding feature importances are shown in fig. 4.4.

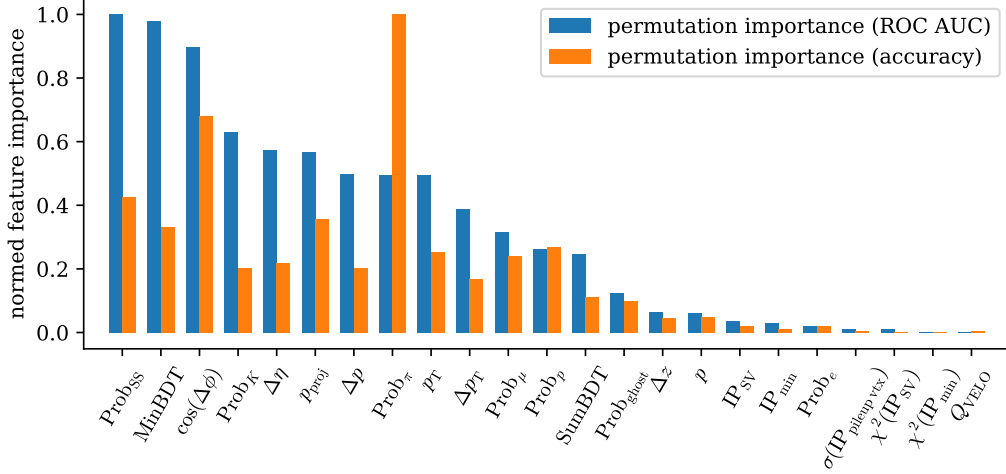


Figure 4.4: Calculated feature importances on the trained DeepSet. Shown are the permutation importances of the ROC AUC and the accuracy normed on the largest value.

Multiple features that are used here are already described in section 4.2. These features are p , p_T , p_{proj} , Δp_T , Δz , $\cos(\Delta\phi)$, $\Delta\eta$, Prob_{ghost}, IP_{SV}, $\chi^2(\text{IP}_{\text{SV}})$, IP_{min}, SumBDT and MinBDT. The other features used here are described in the following paragraph.

All the following features describe a single particle or rather its reconstructed track. Δp is the difference of the absolute momenta of the particle and the signal B . $\chi^2(\text{IP}_{\text{min}})$ is the uncertainty of IP_{min}. Prob_{SS} is an estimated probability that the track belongs to the SS and is the output of the classifier described in section 4.2. Prob_e, Prob_K, Prob _{μ} , Prob_p and Prob _{π} are estimated probabilities about the particle's type. Q_{VELO} is a measure for the interaction of the particle with the VELO detector based on the induced current due to material interactions and bremsstrahlung photons.

The final DeepSet consists of a ϕ -network that extracts latent space vectors of length 64 for each track and a ρ -network which transforms the sum of these vectors to a scalar output. The ϕ -network has an input layer of size 23 followed by 3 layers with of sizes 64, 128 and 64. The ρ -network has an input layer of size 64 followed by 3

Table 4.2: List of all features used to train the DeepSet for B meson classification.

feature	feature
p	Prob _{SS}
p_T	Prob _e
p_{proj}	Prob _{ghost}
Δp	Prob _K
Δp_T	Prob _{μ}
Δz	Prob _p
$\cos(\Delta\phi)$	Prob _{π}
$\Delta\eta$	$\sigma(\text{IP}_{\text{pileup vtx}})$
IP _{SV}	Q_{VELO}
$\chi^2(\text{IP}_{\text{SV}})$	SumBDT
IP _{min}	MinBDT
$\chi^2(\text{IP}_{\text{min}})$	

layers of sizes 128, 64 and 1. All hidden layers use the ReLU activation function and have a dropout rate of 50%. The output layer of the ρ -network uses the sigmoid activation function.

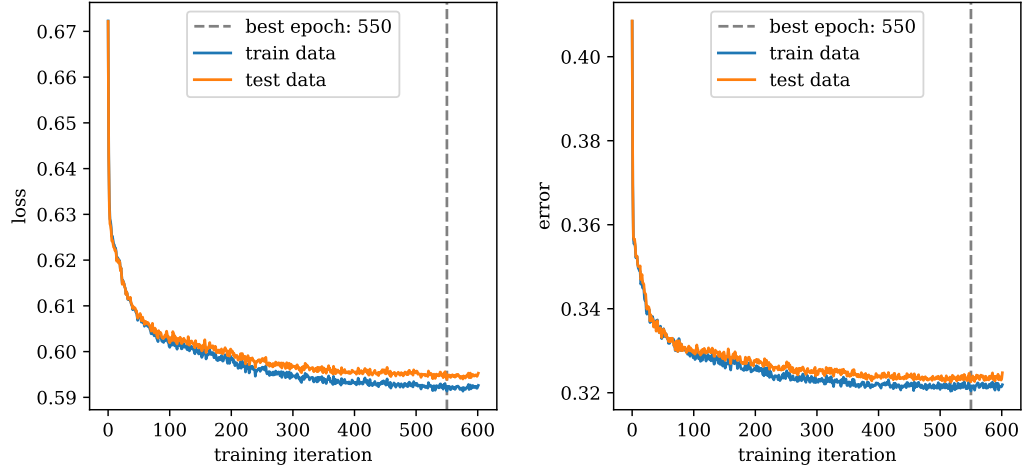
From the 0.4 million events in the dataset, 60% are used for training and 40% are used for validation of the trained model. Before training or evaluation, each feature in the data is rescaled using

$$f(x) = \frac{x - \mu_i}{\sigma_i} \quad (4.1)$$

to ensure a common order of magnitude for all input features. Where μ_i is the mean and σ_i is the standard deviation of feature i in the training data.

The DeepSet is trained using the loss function *binary cross entropy* and the optimizer *Adam*. Using early stopping, the training is stopped at iteration 600 after 50 iterations without improvement on the validation dataset. The loss and the error rate for each iteration of the training of the DeepSet are shown in fig. 4.5.

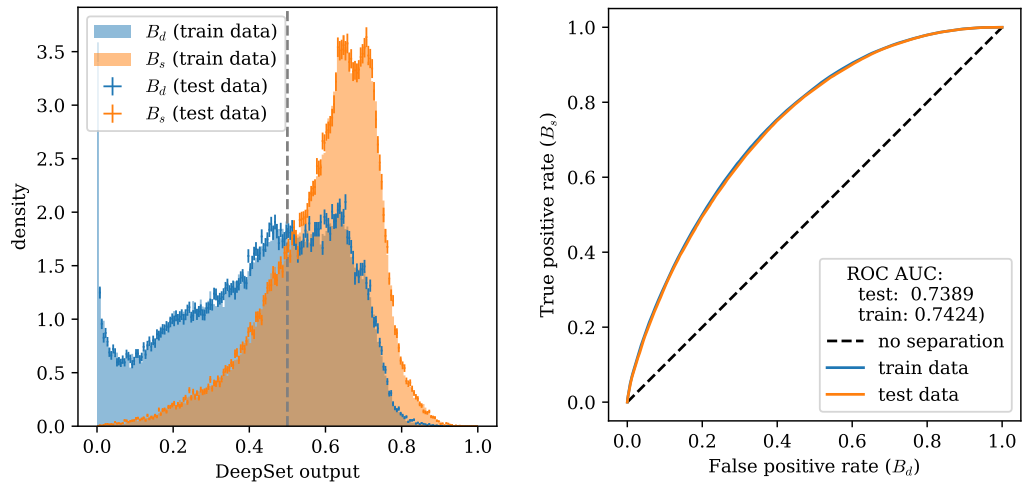
To show the achieved separation of the B mesons, histograms of Prob _{B_s} split by the true labels of the simulation are shown in fig. 4.6a. The achieved ROC curve is shown in fig. 4.6b with a ROC AUC of 0.739 on the test data and 0.742 on the training data. Therefore, a clear separation of the B mesons is achieved on simulation. However, any algorithm that has been developed with simulated samples has to be tested on data that originates from actual measurements. In the following section, the developed classifier is tested on LHCb data of the decays B_d or $B_s \rightarrow J/\psi K_S^0$.



(a) Binary cross entropy loss during training

(b) Error rate during training

Figure 4.5: Performance of the DeepSet for B meson classification after each training iteration.



(a) Distribution of Prob_{B_s}

(b) ROC curve of the DeepSet predictions

Figure 4.6: The left figure shows the distribution of the DeepSet output split by the true labels of the simulation. The right figure shows the ROC curve of the DeepSet output. Both figures show the DeepSet prediction for the test data and the training data.

4.4 Testing the algorithm on LHCb data

Although the developed model of this thesis is intended to distinguish B_d from B_s mesons in pp -collisions independent of the decay channel, it is trained only on simulated samples containing the decay channels $B_d \rightarrow J/\psi K^*$ and $B_s \rightarrow D_s^+ \pi^-$. Consequently, the trained model must be evaluated on LHCb data of some other decay channel. Chosen here are the decays B_d or $B_s \rightarrow J/\psi K_S^0$. The dataset used here contains LHCb data from the years 2016, 2017 and 2018 that has been selected for both mentioned decays using similar selection criteria.

Since this dataset originates from a real measurement rather than a simulation, the data contains a significant amount of background events. In order to be able to fit the B_s resonance, the amount of background events is reduced. The procedure used for background reduction is based on the work-in-progress measurement of $\sin(2\beta)$ in $B^0 \rightarrow J/\psi K_S^0$ decays with run 2 data, and is explained here only very briefly.

A few types of background are reduced manually. One source of such background events is the misidentification of $\Lambda^0 \rightarrow p^+ \pi^-$ decays as $K_S^0 \rightarrow \pi^+ \pi^-$. To reduce this type of background, events are rejected if the reconstructed invariant mass of the K_S^0 candidate is in the region of the Λ^0 mass (1095 MeV to 1140 MeV) and $\text{Prob}_p > 0.10$ is required. Misidentified backgrounds from $B \rightarrow J/\psi K^*$ decays are rejected by requiring that the kaon candidate has a lifetime longer than 0.5 ps. Finally, a cut on the general reconstruction uncertainty of the event is applied ($\chi^2(\text{fit}) < 5.0$).

To reduce all other types of combinatorial background, a BDT is trained. Again, the BDT is implemented in Python using the library XGBoost [12]. For this, events of the mentioned dataset with a reconstructed signal B mass larger than 5450 MeV are used as background events, and simulated events of the decay $B_d \rightarrow J/\psi K_S^0$ are used as signal events. The BDT has a maximum tree depth of 2, is trained at a learning rate of 0.3 and consists of 2000 estimators. A ROC AUC score of 0.989 is achieved on test data and the corresponding ROC curve is shown in fig. 4.7a. With this, events are classified as background events if the BDT output is smaller than 0.5.

The features used by the BDT are listed in table 4.3 and describe the signal B and its decay particles. The symbols inside the brackets represent the corresponding particle in the signal decay. $\text{IP}(x)$ is the impact parameter to the production vertex of particle x . $p_T(x)$ is the transversal momentum of particle x . $\eta(x)$ is the pseudorapidity of particle x . $p_z(x)$ is the momentum of particle x along the z -axis. $\text{FD}(x)$ is the flight distance from the production vertex to the decay vertex

Table 4.3: List of all features used to train the BDT for the distinction between signal and background events.

feature	feature
IP(B^0)	$p_T(\pi^+)$
IP(J/ψ)	$p_T(\pi^-)$
IP(K_S^0)	$p_T(K_S^0)$
IP(μ^+)	$\eta(B^0)$
IP(μ^-)	$\eta(K_S^0)$
FD(K_S^0)	$p_z(K_S^0)$
$\chi^2(\text{fit})$	

of particle x . Finally, $\chi^2(\text{fit})$ is the χ^2 corresponding to the reconstruction of the entire signal decay.

Figure 4.7b shows the distribution of the reconstructed signal B mass before and after reducing the background with the BDT and the manual cuts. After reducing the background, the trained model for the signal B classification is applied on the data to calculate Prob_{B_s} . To estimate the performance of this prediction, an alternative method to estimate the amount of B_d and B_s mesons is needed. The chosen procedure is to do a least squares fit on the B mass distribution and estimate the counts of events by integrating each component of the fitted function.

In the following function definitions every variable except M_B is a parameter that has to be fitted. The B mass distribution found in the data can be approximated by a function

$$F(M_B) = N_{\text{bkg}} \cdot F_{\text{bkg}}(M_B) + N_{B_d} \cdot F_{B_d}(M_B) + N_{B_s} \cdot F_{B_s}(M_B) \quad (4.2)$$

that is split into three components. The first component F_{bkg} describes the combinatorial background with

$$F_{\text{bkg}}(M_B) = \exp(-\lambda \cdot M_B). \quad (4.3)$$

The second component F_{B_d} describes the events containing a B_d meson and the third component F_{B_s} describes the events containing a B_s meson. Both F_{B_d} and F_{B_s} use the same function

$$\begin{aligned} F_B(M_B) = & f_1 \cdot f_2 \cdot F_{\text{CB}}\left(\frac{M_B - \mu}{\sigma_1}, \beta_1, m_1\right) \\ & + (1 - f_1) \cdot f_2 \cdot F_{\text{CB}}\left(-\frac{M_B - \mu}{\sigma_2}, \beta_2, m_2\right) \\ & + (1 - f_1) \cdot (1 - f_2) \cdot F_{\text{gauss}}(M_B, \mu, \sigma_3), \end{aligned} \quad (4.4)$$

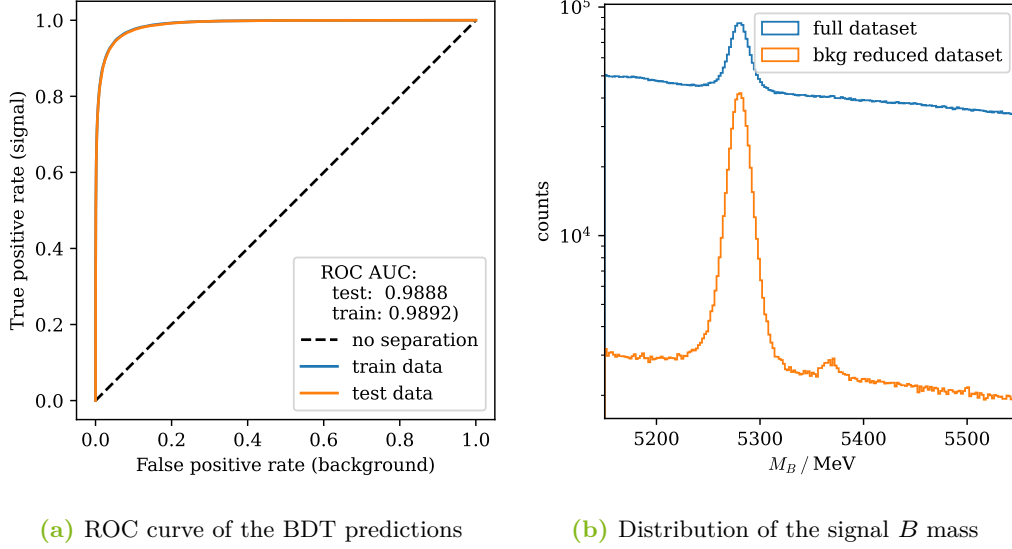


Figure 4.7: The left figure shows the ROC curve achieved by the BDT for background identification. The right figure shows the distribution of the reconstructed signal B mass before and after reducing the background with the BDT and the manual cuts.

with the same parameters except $\mu_{B_s} = \mu_{B_d} + (M_{B_s} - M_{B_d})$. $M_{B_d} = 5279.65$ MeV is the mass of B_d mesons and $M_{B_s} = 5366.88$ MeV is the mass of B_s mesons [14].

$$F_{CB}(x, \beta, m) = \begin{cases} N \cdot \exp(-\frac{x^2}{2}) & \text{for } x > -\beta \\ N \cdot \left(\frac{m}{|\beta|}\right)^m \cdot \exp\left(-\frac{\beta^2}{2}\right) \cdot \left(\frac{m}{|\beta|} - |b| - x\right)^{-m} & \text{for } x \leq -\beta \end{cases} \quad (4.5)$$

is called a Crystal Ball distribution and

$$F_{\text{gauss}}(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) \quad (4.6)$$

is called a normal distribution.

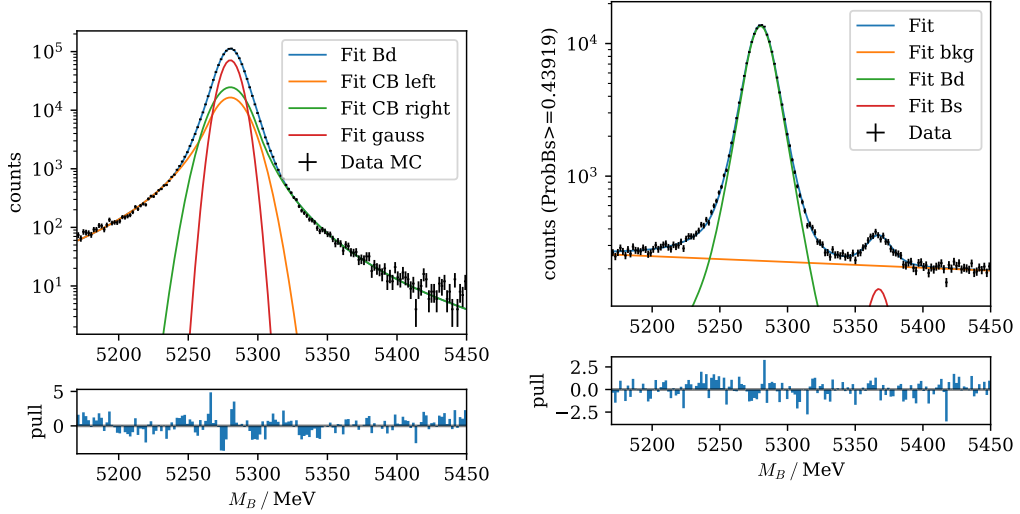
All the following fits are done using the library *iminuit*[15]. First, a fit is done on a simulated sample of the B_d mode to fix the parameters $f_1, f_2, \beta_1, \beta_2, m_1$ and m_2 for all the following fits. This fit is shown in fig. 4.8a. Then, for 100 values of x between 0 and 1, a selection with $\text{Prob}_{B_s} \leq x$ and a selection with $\text{Prob}_{B_s} \geq x$ is

made. A fit is done on each acquired B mass distribution and the counts

$$n_{B_d} = \int_{5170 \text{ MeV}}^{5450 \text{ MeV}} N_{B_d} \cdot F_{B_d}(M_B) dM_B \quad \text{and} \quad (4.7)$$

$$n_{B_s} = \int_{5170 \text{ MeV}}^{5450 \text{ MeV}} N_{B_s} \cdot F_{B_s}(M_B) dM_B \quad (4.8)$$

are calculated numerically. One of those fitted distributions is shown in fig. 4.8b.



(a) Fit of the simulated B_d mass distribution (b) Example fit of the B mass distribution after a cut on LHCb data

Figure 4.8: The left figure shows a fit on a simulated B_d sample. The right figure shows a fit of the B mass distribution on LHCb data with the selection $\text{Prob}_{B_s} \geq 0.43919$. Both figures also show a pull distribution between the data points and the fit.

To estimate how well the trained model of this thesis can distinguish B_d and B_s mesons, the ratio n_{B_s}/n_{B_d} is plotted by the cut value x for selections with $\text{Prob}_{B_s} \geq x$. This is shown in fig. 4.9a. The expected value without any achieved separation is

$$\frac{\text{BR}(B_s \rightarrow J/\psi K_S^0)}{\text{BR}(B_d \rightarrow J/\psi K_S^0)} \cdot f_s/f_d(13 \text{ TeV}) = 0.0109 \pm 0.0010. \quad (4.9)$$

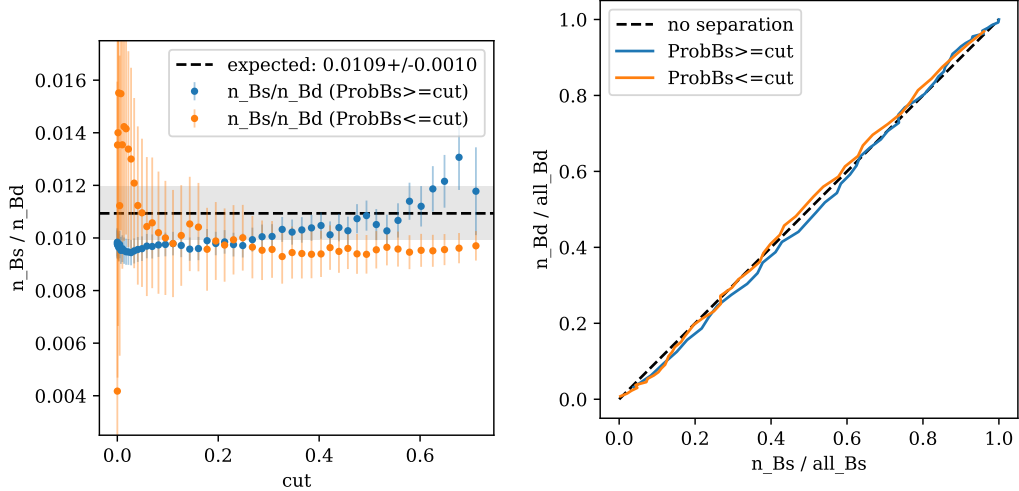
Where the branching ratios are $\text{BR}(B_s \rightarrow J/\psi K_S^0) = (4.37 \pm 0.16) \times 10^{-4}$ and $\text{BR}(B_d \rightarrow J/\psi K_S^0) = (1.88 \pm 0.15) \times 10^{-5}$ [14], and $f_s/f_d(13 \text{ TeV}) = 0.2539 \pm 0.0079$

is the ratio of the B_s^0 and B_d^0 fragmentation fractions in pp -collisions at 13 TeV collision energy [16].

Another approach to show the achieved separation, is to plot a curve similar to a ROC curve. For each selection, the estimated efficiencies

$$\varepsilon_{B_d} = \frac{n_{B_d}(x)}{n_{B_d}(\text{no cut})} \quad \text{and} \quad \varepsilon_{B_s} = \frac{n_{B_s}(x)}{n_{B_s}(\text{no cut})} \quad (4.10)$$

are calculated. Shown in fig. 4.9b is the plot of ε_{B_d} against ε_{B_s} . Similar to a ROC curve, a diagonal line from (0,0) to (1,1) is the expectation without any achieved separation. Compared to the achieved ROC curve on a simulation shown in fig. 4.6b, a substantially smaller separation is achieved on LHCb data of the decays B_d or $B_s \rightarrow J/\psi K_S^0$.



(a) Ratio n_{B_s}/n_{B_d} by cut value on Prob_{B_s} (b) Efficiencies of the Prob_{B_s} selections similar to a ROC curve

Figure 4.9: The left figure shows the estimated ratio n_{B_s}/n_{B_d} after applying Prob_{B_s} cuts on LHCb data. Here, it is important to note that selections with $\text{Prob}_{B_s} \leq x$ on small x have a very small B_s peak, which makes the interpretation of these ratios almost impossible. The right figure shows the estimated efficiencies ε_{B_d} against ε_{B_s} after applying Prob_{B_s} cuts on LHCb data. In both figures, selections with $\text{Prob}_{B_s} \geq x$ are shown in blue and selections with $\text{Prob}_{B_s} \leq x$ are shown in orange.

5 Conclusion

In this thesis, a classification algorithm to distinguish neutral B_d from neutral B_s mesons in pp -collisions at LHCb is developed and trained on simulation and evaluated on LHCb data. This algorithm is intended to be independent of the signal B decay channel and uses only data of the tracks associated with the signal B meson without tracks of the signal decay. For the training of supervised machine learning algorithms, simulated samples of the decays $B_d \rightarrow J/\psi K^*$ and $B_s \rightarrow D_s^+ \pi^-$ are used.

A Boosted Decision Tree (BDT) classifier is trained to identify *same side* (SS) fragmentation tracks. The SS fragmentation contains particles that are related to the $d\bar{d}$ - or $s\bar{s}$ -pair that is part of the production of the signal B . In contrast to other tracks in the data SS tracks therefore contain relevant information for the distinction between B_d and B_s mesons. The output of this BDT is then used as an additional input for a subsequent DeepSet model for the signal B classification.

A BDT is trained using all available features and features are then selected based on feature importance metrics called the gain and the permutation importance of the ROC AUC. The final BDT uses 21 features, and a ROC AUC of 0.763 is achieved on test samples. Although the achieved separation is not sufficient to correctly identify most of the SS tracks, this BDT output should help the DeepSet to identify relevant information.

For the feature selection of the DeepSet, the permutation importances of the accuracy and the ROC AUC are used to select a set of 23 training features with which the final DeepSet classifier is trained. An indication that the BDT output (Prob_{SS}) has an impact on the DeepSet performance is observed. This however has to be interpreted with care, because feature importances can only be seen as estimates. The trained DeepSet achieves a ROC AUC of 0.739 on test data indicating a clear separation.

Compared to the training data, LHCb data contains background events that do not actually involve B_d or B_s mesons. Since evaluating the B classification requires clearly visible B_s peaks in the B mass distribution, the number of background events has to be reduced beforehand. This background reduction is done using a combination of manual cuts and a BDT that is trained on simulated samples of the decay $B_d \rightarrow J/\psi K_S^0$.

The developed algorithm of this thesis is then applied to the background reduced data. To evaluate the achieved separation, a model is fitted to the B mass distributions after various cuts on the DeepSet output (Prob_{B_s}) are applied. The fit components

that represent the combinatorial background, the B_d peak and the B_s peak are then integrated numerically. This results in estimated counts of events attributed to the B_d and B_s mesons remaining after each selection.

The resulting ratios of the number of B_s decays over the number of B_d decays are calculated. Here, selection criteria of the form $\text{Prob}_{B_s} \geq x$ show an upward trend with increasing x values. This could be an indication of some level of achieved separation between B_s and B_d mesons. Selection criteria of the form $\text{Prob}_{B_s} \leq x$ show an increase of this ratio for lower x values. However, for small values of x in $\text{Prob}_{B_s} \leq x$ there are too few events left and an estimation of the B_s peak becomes increasingly difficult. This is not the case for large values of x in $\text{Prob}_{B_s} \geq x$ where a clear B_s peak is visible. All calculated ratios are compatible with the expected ratio for no achieved separation, but the upward trend for $\text{Prob}_{B_s} \geq x$ selections suggests at least some level of achieved separation. Another method to test the achieved separation is to plot the efficiencies of both the B_d and B_s mesons in the form of a ROC curve. Here, no clear signs of an achieved separation are visible.

In summary, the classification algorithm for separating B_s and B_d mesons is designed and tested on simulation and LHCb data. Unfortunately, the successful separation of B_s and B_d mesons on simulated samples could not be reproduced on LHCb data. However, the achieved separation on simulation is an indication that the strategy to use a DeepSet to distinguish B_d from B_s mesons based on data of the associated event could succeed. The observed performance loss raises the question whether the classifier exploited selection differences in the training datasets that went unnoticed, or if a simulated feature is mismodeled. In a future analysis it will therefore be important to show that observed kinematic differences in the training simulation are only due to differences in the B_d and B_s masses and that their distributions closely match the data. Since it cannot be decided with absolute certainty that kinematic differences in the training simulation are only due to B_d - B_s mass differences and not due to indirect selection constraints, an attempt could be made at equalizing the kinematics in the training data through reweighting or similar methods. If successful, this study can be extended to include charged B modes.

Bibliography

- [1] Super-Kamiokande Collaboration, Y. Fukuda et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.81.1562>.
- [2] SNO Collaboration, Q. R. Ahmad et al. “Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory”. In: *Phys. Rev. Lett.* 89 (2002), p. 011301. DOI: 10.1103/PhysRevLett.89.011301. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.89.011301>.
- [3] LHCb Collaboration, R. Aaij et al. “Observation of the Resonant Character of the $Z(4430)^-$ State”. In: *Phys. Rev. Lett.* 112 (2014), p. 222002. DOI: 10.1103/PhysRevLett.112.222002. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.112.222002>.
- [4] LHCb Collaboration, R. Aaij et al. “Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \rightarrow J/\psi K^- p$ Decays”. In: *Phys. Rev. Lett.* 115 (2015), p. 072001. DOI: 10.1103/PhysRevLett.115.072001. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.115.072001>.
- [5] ATLAS Collaboration, G. Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2012.08.020>. URL: <https://www.sciencedirect.com/science/article/pii/S037026931200857X>.
- [6] CMS Collaboration, S. Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2012.08.021>. URL: <https://www.sciencedirect.com/science/article/pii/S0370269312008581>.
- [7] U. Eitschberger. “Flavour Tagging Plots for Conference”. 2016. URL: https://twiki.cern.ch/twiki/bin/view/LHCb/FlavourTaggingConferencePlots#Flavour_Tagging_Plots_for_Confer (visited on 06/14/2022).
- [8] D. P. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. In: (2014). DOI: 10.48550/ARXIV.1412.6980. URL: <https://arxiv.org/abs/1412.6980>.
- [9] M. Zaheer et al. “Deep Sets”. In: (2017). DOI: 10.48550/ARXIV.1703.06114. URL: <https://arxiv.org/abs/1703.06114>.

-
- [10] L. Evans and P. Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001. URL: <https://doi.org/10.1088/1748-0221/3/08/s08001>.
 - [11] LHCb Collaboration, A. Augusto Alves et al. “The LHCb Detector at the LHC”. In: *Journal of Instrumentation* 3.08 (2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005. URL: <https://doi.org/10.1088/1748-0221/3/08/s08005>.
 - [12] T. Chen and C. Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. URL: <http://doi.acm.org/10.1145/2939672.2939785>.
 - [13] A. Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035. URL: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
 - [14] Particle Data Group, P. A. Zyla et al. “Review of Particle Physics”. In: *PTEP* 2020.8 (2020). and 2021 update, p. 083C01. DOI: 10.1093/ptep/ptaa104.
 - [15] H. Dembinski, P. Ongmongkolkul et al. “scikit-hep/iminuit”. In: (2020). DOI: 10.5281/zenodo.3949207. URL: <https://doi.org/10.5281/zenodo.3949207>.
 - [16] LCHb Collaboration, R. Aaij et al. “Precise measurement of the f_s/f_d ratio of fragmentation fractions and of B_s^0 decay branching fractions”. In: *Phys. Rev. D* 104 (2021), p. 032005. DOI: 10.1103/PhysRevD.104.032005. URL: <https://link.aps.org/doi/10.1103/PhysRevD.104.032005>.

Danksagung (Acknowledgements)

In erster Linie möchte ich mich herzlich bei Herrn Prof. Dr. Albrecht bedanken, dass ich die Möglichkeit erhalten habe, ein Teil seiner Arbeitsgruppe zu sein und meine Bachelorarbeit dort durchzuführen. Weiter bedanke ich mich bei Herrn Dr. Elsässer, dass er sich bereit erklärt hat, der Zweitgutachter dieser Bachelorarbeit zu sein.

Mein außerordentlicher Dank gebührt meinen Betreuern Vukan und Quentin, die mich über die gesamte Zeit der Bachelorarbeit unterstützten und mir auch außerhalb von gewöhnlichen Arbeitszeiten sehr viele Ratschläge gegeben haben. Zusätzlich bedanke ich mich bei Vukan für den großen Zeitaufwand in der Endphase der Bachelorarbeit. Außerdem vielen Dank an Sophie für das gründliche Korrekturlesen dieser Arbeit und die vielen hilfreichen Anmerkungen.

Allgemein möchte ich mich bei allen Mitgliedern der Arbeitsgruppe Albrecht für das fantastische Arbeitsklima bedanken. Zudem wertschätze ich alle Hinweise und Ratschläge, die mir persönlich oder unter anderem im wöchentlichen CPV-Meeting gegeben wurden.

Abschließend bedanke ich mich bei meiner Familie und meinen Freunden für die liebevolle Unterstützung während meines gesamten Bachelorstudiums.

Eidesstattliche Versicherung

(Affidavit)

Guth, Nico

Name, Vorname
(surname, first name)

205945

Matrikelnummer
(student ID number)

☒ Bachelorarbeit
(Bachelor's thesis)

☐ Masterarbeit
(Master's thesis)

Titel
(Title)

Development of a multivariate algorithm for the classification of
B mesons at the LHCb experiment

Ich versichere hiermit an Eides statt, dass ich die vorliegende Abschlussarbeit mit dem oben genannten Titel selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

I declare in lieu of oath that I have completed the present thesis with the above-mentioned title independently and without any unauthorized assistance. I have not used any other sources or aids than the ones listed and have documented quotations and paraphrases as such. The thesis in its current or similar version has not been submitted to an auditing institution before.

Dortmund, 14.06.2022

Ort, Datum
(place, date)


Unterschrift
(signature)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -).

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird ggf. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offense can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offenses of this type is the Chancellor of TU Dortmund University. In the case of multiple or other serious attempts at deception, the examinee can also be unenrolled, Section 63 (5) North Rhine-Westphalia Higher Education Act (*Hochschulgesetz, HG*).

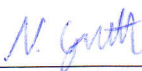
The submission of a false affidavit will be punished with a prison sentence of up to three years or a fine.

As may be necessary, TU Dortmund University will make use of electronic plagiarism-prevention tools (e.g. the "turnitin" service) in order to monitor violations during the examination procedures.

I have taken note of the above official notification:*

Dortmund, 14.06.2022

Ort, Datum
(place, date)


Unterschrift
(signature)

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the Bachelor's/ Master's thesis is the official and legally binding version.