In unserer Arbeit haben wir zum einen die Funktionsweise der Support Vector Machines als binäre Klassifikationsmethode beleuchtet, als auch die Leistungsfähigkeit in Datenszenarien mit verschiedenen Eigenschaften verglichen. Diese Datensszenarien, wurden synthetisch von uns hergestellt und sind in ihrem datengenerierenden Prozess speziell auf *SVM* zugeschnitten. Anschließend haben wir uns mit einschlägiger Literatur beschäftigt, bei der die Autoren bereits ähnliche Versuche durchgeführt haben und anhand dessen Hypothesen abgeleitet. Für die Durchführung unserer Versuche haben wir die Performance von *SVM* mit verschiedenen Kernels sowie weiterer Klassifikationsmethoden über alle Szenarien mit verschiedenen Maßzahlen verglichen. Insgesammt mussten wir allerdings festellen, dass unsere ergebnisse die Hypothesn nur in Teilen bestätigen. *SVM*-Methoden, haben im Durschnitt oft bessere Leistungen gezeigt, als die anderen Methoden. Wir schließen aufgrund der Rankings, dass, egal welche Dimesnionierung vorliegt und wenn keine Information über die Form der Entscheidungsgrenze vorliegt, *SVM* mit radialen und polynomialen Kernel immer eine gute Entscheidung darstellen. Zumindest sollten diese beiden dem *SVM* mit linearem Kernel vorgezogen werden. Wobei hier auch anzumerken ist, dass mit der höheren Flexibilität der *SVM-P* und *SVM-R* auch die Gefahr besteht, dass es zu Overfitting kommt. Dies könnte vor allem bei empirischen Daten zu höheren Klassifikationsfehlern führen.
Allerdings haben wider Erwarten die SVM mit dem Kernel der auf die Datenerzeugung eigentlich zugeschnitten ist nicht besser performt. Wir vermuten, dass es an einem zu niedrigen n-Wert in der Dimensionierung für die Szenarien 4 bis 9 liegen könnte und dass sich in einer etwaigen Simulationsstudie mit dem gleichen DGP und Dimensionierung die Hypothese vielleicht doch noch bestätigen könnte, da sich so zufällige Abweichungen aufheben. Was das Verhalten in den verschiedenen Dimensionalitäten angeht können wir zumindest sagen, dass in Fällen mit mehr Beobachtungen als Variablen *SVM* eine gute Wahl darstellen. Schwieriger wird es bei den hochdimensionalen Szenarien, da hier *K-NN* eindeutig besser performt hat. Uns fällt es schwer dies zu erklären, da wir eigentlich davon ausgingen, dass *K-NN* bei vielen Variablen eher schlechter performt.

Wir sehen daher noch einige Optimierungsmöglichkeiten für weitere Arbeiten dieser Art. So wäre es angebracht, wie bereits gesagt die Datengererierung für die einzelnen Szenarien wiederholt durchzuführen und die Ergebnisse zu mitteln. Die Dimesnionierung könnte auch so angepasst werden, dass die n Werte etwas höher sind auch für Szenarien mit $p \gg n$.
Auch könnte der Einfluss der Szenarien auf die Berechnungszeit für die *SVM*-Algorithmen noch einbezogen werden. Ein Benchmark-Test könnte auch hier zu interessanten Ergebnissen führen. Zusätzlich haben wir hier lediglich eine handvoll Klassifikationsalgorithmen im Vergleich untersucht. Eine Erweiterung auf *classification Trees,Discriminant Analyses* oder verschiedene Ensemble-Methoden ist denkbar. An der Datengenerierung liesen sich auch noch weitere Aspekte anpassen. So könnte die Anteile der Ausprägungen in der binären Zielvariable noch variiert, mehr als zwei Ausprägungen generiert oder auch noch komplexere Entscheidungsgrenzen modelliert werden. Diese Erweiterungen hätten allerdings den Rahmen dieser Arbeit überschritten. Trotzdem ergänzt unsere Arbeit die bisherigen Befundem zur Leistungs- und Anpassungsfähigkeit von Support-Vector Machines in verschiedenen Datensituation, sowie deren Bedeutung im Kontext von Klassifikationsaufgaben.