**Algorithmic Composition of Classical Music through Data Mining**

An All College Thesis

College of Saint Benedict and Saint John’s University

by

Tom Donald Richmond

April, 2018

Algorithmic Composition of Classical Music through Data Mining

By Tom Donald Richmond

**Approved By**:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dr. Imad Rahal, Professor of Computer Science

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dr. Mike Heroux, Scientist-in-Residence of Computer Science

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dr. Jeremy Iverson, Assistant Professor of Computer Science

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dr. Imad Rahal, Chair, Computer Science Department

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Molly Ewing, Director, All-College Thesis Program

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Jim Parsons, Director, All-College Thesis Program

**Abstract**

The desire to teach a computer how to algorithmically compose music has been a topic in the world of computer science since the 1950’s, with roots of computer-less algorithmic composition dating back to Mozart himself. One limitation of algorithmically composing music has been the difficulty of eliminating the human intervention required to achieve a musically homogenous composition. We attempt to remedy this issue by teaching a computer how the rules of composition differ between the six distinct eras of classical music by having it examine a dataset of musical scores, rather than explicitly telling the computer the formal rules of composition. To pursue this an automated composition process, we examined the intersectionality of algorithmic composition with the machine learning concept of classification. Using a Naïve Bayes classifier system, the computer classifies pieces of classical music into their respective era based upon a number of attributes. It attempts to recreate each of the six classical styles using a technique inspired by cellular automata. The success of this process is twofold determined by feeding composition samples into a number of classifiers, as well as analysis by studied musicians. We concluded that there is potential for further hybridization of classification and composition techniques.

**Table of Contents**

**1. Introduction …………………………………………………………………… 1**

1.1 Early Explorations ……………………………………………………. 1

1.2 The Data-Driven Intelligence Age ……………………………………. 2

1.3 Study Overview ………………………………………………………. 3

**2. Data ……………………………………………………………………………. 4**

2.1 Musical Representations …………………………………………….... 4

2.2 Digital Formats ……………………………………………………….. 5

2.2.1 MIDI ………………………………………………………… 6

2.2.2 \*\*kern ………………………………………………………. 6

**3. Data Mining …………………………………………………………………… 7**

3.1 Data Extraction ……………………………………………………….. 8

3.1.1 Classes ……………………………………………………… 8

3.1.2 Attributes …………………………………………………… 10

3.1.3 Pre-Process …………………………………………………. 12

3.2 Classification …………………………………………………………. 13

3.3 Results ………………………………………………………………… 15

**4. Generation …………………………………………………………………….. 16**

4.1 Methods ………………………………………………………………. 17

4.1.1 Cellular Automata …………………………………………... 17

4.1.2 Adapted Musical Model …………………………………….. 18

4.2 Results Analysis ……………………………….……………………… 21

4.2.1 Machine Analysis …………………….…..…………………. 22

4.2.2 Expert Analysis ……………………………………………… 22

**5. Discussion ……………………………………………………………………… 23**

5.1 Conclusion ……………………………..……………………………… 24

5.2 Applications …………………………………………………………… 24

5.3 Future Works ………………………………………………………….. 25

**6. Appendix ...…………………………………………………………………….. 26**

**7. References ……………………………………………………………………... 46**

**List of Figures**

Figure 1 - Timeline of musical eras ……………………………………………… 9

Figure 2 - List of Attributes ……………………………………………………… 10

Figure 3 - Chromatic Circle ……………………………………………………… 11

Figure 4 - Image of intervals ……………………………………………………... 12

Figure 5 - Data set and distribution chart ………………………………………… 13

Figure 6 - Chart of classification results ………….……………………………… 15

Figure 7 - Image of Wolfram Algorithm and a rule set ………………………….. 18

Figure 8 - Mapping of binary sequences to notes ………………………………… 19

Figure 9 - Sample extraction of Naïve Bayes ……………………..……………… 20

Figure 10 - Example of transitionary rule mapping ………………………………. 20

Figure 11 - Chart of generation results ………………………………….………. 22