

# Contents

Abstract	v
Resumen	vii
Resum	ix
Acknowledgements	xi
Contents	xiii
List of Figures	xvii
List of Tables	xix
<b>I Introduction</b>	<b>1</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Motivation . . . . .	3
1.2 Research Questions and Objectives . . . . .	6
1.3 Research Methodology . . . . .	7
1.4 Thesis Outline . . . . .	9
<b>II Background</b>	<b>15</b>
<b>2 Inductive Programming</b>	<b>17</b>
2.1 Introduction . . . . .	17
2.2 Learning programs . . . . .	21
2.3 Inductive Bias . . . . .	23
2.4 General Purpose Inductive Programming Systems . . . . .	25
<b>3 Data Science Automation</b>	<b>31</b>
3.1 Introduction . . . . .	31
3.2 Data Science Trajectories . . . . .	33
3.3 Data Wrangling . . . . .	35
3.4 Data Wrangling Automation . . . . .	38
	xiii

## Contents

---

<b>III</b>	<b>BK-ADAPT: Automating Data Format Standardisation</b>	<b>43</b>
<b>4</b>	<b>Domain-specific Induction</b>	<b>45</b>
4.1	Introduction . . . . .	45
4.2	Problem Definition . . . . .	48
4.3	Experiments . . . . .	52
4.4	Conclusions . . . . .	55
<b>5</b>	<b>Dynamic Background Knowledge</b>	<b>59</b>
5.1	Introduction . . . . .	59
5.2	Upgraded Approach . . . . .	60
5.3	Method . . . . .	61
5.4	Experiments . . . . .	63
5.5	Conclusions . . . . .	70
<b>IV</b>	<b>AUTOMAT[R]IX: Automating Matrix Transformations</b>	<b>73</b>
<b>6</b>	<b>Learning Simple Matrix Pipelines</b>	<b>75</b>
6.1	Introduction . . . . .	76
6.2	Problem Definition . . . . .	78
6.3	Method . . . . .	79
6.4	Experiments . . . . .	84
6.5	Conclusions . . . . .	93
<b>V</b>	<b>Conclusions</b>	<b>95</b>
<b>7</b>	<b>Conclusions and Future Work</b>	<b>97</b>
7.1	Conclusions . . . . .	97
7.2	Future Work . . . . .	100
	<b>Bibliography</b>	<b>103</b>
	<b>Appendices</b>	<b>117</b>
<b>A</b>	<b>Data Collection</b>	<b>119</b>
A.1	Survey to collect personal data . . . . .	119
<b>B</b>	<b>Data Wrangling: Competences and Skills</b>	<b>121</b>
B.1	List of Data Wrangling Competences . . . . .	122
B.2	List of Data Wrangling Hard Skills . . . . .	122
B.3	List of Data Wrangling Soft Skills . . . . .	123
<b>C</b>	<b>Logging Data Scientists</b>	<b>125</b>

<b>D</b>	<b>BK-ADAPT: Supplementary Material</b>	<b>129</b>
D.1	Background Knowledge: List of Functions . . . . .	129
D.2	Data: List of Meta-features . . . . .	140
D.3	Experiments: Extended Results . . . . .	142
D.4	Tool: System Overview . . . . .	149
<b>E</b>	<b>AUTOMAT[R]IX : Supplementary Material</b>	<b>153</b>
E.1	Background Knowledge: List of Functions . . . . .	153
E.2	Prior Probabilities: List of Top Functions in GitHub . . . . .	155
E.3	Text Hints: Frequent Terms from Function’s Documentation . . . . .	156
E.4	Data: List of Examples . . . . .	160
E.5	Pipeline of events for a simple example . . . . .	161