**Chapter 3**

**Research problem and statement of objectives**

**Problem Statement**

Emojis have undoubtedly become a ubiquitous part of our online lives, first appearing in 1999 and their use becoming increasingly commonplace throughout digital communication. Their usage has resulted in an unprecedented evolution in how digital communication occurs, seamlessly integrating images into typography. While natural language processing is a widely utilised and powerful tool in data analytics, contextual understanding, and ambiguity in text still present challenges to these models. Sarcasm is one such source of ambiguity which has been suggested to account for significant limitations in these models, largely due to lack of contextual clues due to the short-form nature of online text; the models cannot know that the implied meaning of the text opposes the literal. Emojis are widely used for the purpose of establishing tone and clues towards sentiment to the reader. This is intuitive to anyone who consumes digital content, however in the domain of natural language processing, there are few instances of research into the potential for consideration of emojis as a pragmatic indicator in text processing, however literature suggests that this approach has potential. The research problem can be summarised by the following question:

*Can the consideration of pragmatic cues from emojis in text improve outcomes for sarcasm detection?*

**Problem Definition**

The research problem proposed for this work can be clarified using the problem definition model as follows:

*Problem Identification:* Figurative literary devices such as sarcasm present significant challenges to modern natural language processing models as their identification relies heavily on tone, and context including interlocutor relationships which are not captured in the models. Emojis are often used to convey tone within text, however literature exploring their contribution to sarcasm detection is sparse.

*Problem Clarification:* Lexical study of emojis in sarcastic text may be a key avenue for exploration to improve upon present state-of-the-art sarcasm detection for natural language processing.

*Problem Formulation:* Evaluation of emoji pragmatics in sarcastic content compared to literal counterparts may provide additional insight into the tone or context of online content. Integration of these patterns may increase the capacity of natural language processing models to consider these complex linguistic devices for more accurate sentiment analysis outcomes.

**Research Objectives**

To address the overarching research question, the following objectives are proposed:

*Objective 1:* Determine optimal emotion classification methodologies with respect to psychological principles and their implementation within a data analytics context for text and emoji classification. Propose feature extraction strategies for emojis which supplement limitations of pragmatic information in text-based analysis.

*Objective 2:* Establish patterns between text and emoji features which characterise sarcasm in online text. Propose a novel methodology for sarcasm detection which utilises emojis to improve pragmatic understanding of text.

*Objective 3:* Evaluate the impact of integrating emoji pragmatic information into sarcasm detection models. Propose implementations of this novel methodology within real-world natural language processing applications and critically assess its impact.

**Delimitation**

An intuitive delimitation to this work relates to the problem definition which dictates that the conclusions drawn will not apply where emojis are not used. The scope of the data used in this work was limited to tweets generated by native or near-native English speakers. While sarcasm is a universal concept worldwide, conclusions drawn on how it is conveyed within text can only apply to the English-speaking population as this work does not consider the possibility of dynamic semiotics of emojis across languages or cultures. The dataset also limits the scope of any conclusions drawn to twitter data as this work does not assess the consistency in pragmatic patterns across platforms, which may be relevant due to varying limitations and conventions relating to content length, structure and formality of language used.