UNIVERSITY OF CAPE TOWN

MASTERS THESIS

A semantic Bayesian network for automated share evaluation on the JSE

Author:

Rachel DRAKE

Supervisor:

A/Prof. Deshen MOODLEY

Co-supervisor:

Prof. Thomas MEYER

A thesis submitted in fulfillment of the requirements for the degree of Master of Philosophy in Information Technology in the

Department of Computer Science

June 14, 2020

Declaration of Authorship

I, Rachel Katherine Thomson, declare that:

(i) The research reported in this dissertation, except where otherwise indicated, is my original research
(ii) This dissertation has not been submitted for any degree or examination at any other university.
(iii) This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
(iv) This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
   a. their words have been re-written but the general information attributed to them has been referenced.
   b. where their exact words have been used, their writing has been placed inside quotation marks and referenced.
(v) This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.
(vi) I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: ………………

Date: 14th June 2020

Abstract

Advances in information technology have presented the potential to automate investment decision making processes. This will alleviate the need for manual analysis and reduce the subjective nature of investment decision making. However, there are different investment approaches and perspectives for investing which makes acquiring and representing expert knowledge for share evaluation challenging. Current decision models often do not reflect the real investment decision making process used by the broader investment community or may not be well-grounded in established investment theory. This research investigates the efficacy of using ontologies and Bayesian networks for automating share evaluation on the JSE. The knowledge acquired from an analysis of the investment domain and the decision-making process for a value investing approach was represented in an ontology. A Bayesian network was constructed based on the concepts outlined in the ontology for automatic share evaluation. The Bayesian network allows decision makers to predict future share performance and provides an investment recommendation for a specific share. The decision model was designed, refined and evaluated through an analysis of the literature on value investing theory and consultation with expert investment professionals. The performance of the decision model was validated through back testing and measured using return and risk-adjusted return measures. The model was found to provide superior returns and risk-adjusted returns for the evaluation period from 2012 to 2018 when compared to selected benchmark indices of the JSE. The result is a concrete share evaluation model grounded in investing theory and validated by investment experts that may be employed, with small modifications, in the field of value investing to identify shares with a higher probability of positive risk-adjusted returns.

Key Words: Ontology, Bayesian Network, Portfolio Management, Share Evaluation, Value Investing, Fundamental Analysis

Table of Contents

Table of Figures

Glossary of Terms

Term	Meaning
Share/Stock	A unit of ownership in a company which represents one of the equal parts into which a company's capital is divided and a portion of the company's value.
Portfolio	A collection of shares and/or other financial assets held by an individual or organization, which is constructed in accordance with investment objectives.
Stock Exchange	The entity which facilitates buyers and sellers to negotiate prices and make trades on listed securities. Securities traded on a stock exchange include stock issued by listed companies, unit trusts, derivatives, pooled investment products and bonds.
Stock Market	A general term which refers to a place where stocks are traded and refers to the network of exchanges where stocks may be traded.
JSE (Johannesburg Stock Exchange)	The main exchange in South Africa and is the largest stock exchange in Africa.
Return	The profit (or loss) from an investment.
Share Price Performance	The increase (or decrease) in the market price of a share.
Share Evaluation	The evaluation of a particular share to determine the expected future share price performance.
Active Investment Strategy	An investment strategy that involves ongoing buying and selling of shares by an investor. The strategy involves continual monitoring of share activity to exploit profitable conditions.
Value Investing	An investment strategy that involves picking stocks that appear to be trading for less than their intrinsic value determined through some form of fundamental analysis.
Share Index	A share index, or stock market index, is an index that measures a stock market or a subset of the stock market, and allows investors to compare a current price levels with past prices to calculate market performance. The share index is generally computed from the prices of the selected stocks.

Acknowledgements

The effort and care taken with this dissertation are a direct result of my supervisor, Deshen Moodley, who carefully guided and encouraged me to learn and evolve throughout the research process. Further acknowledgement goes to Prof. Tommie Meyer who provided sound input at important stages of the research journey. Lastly, thanks to my husband who has supported me throughout.

1. Introduction

Digital data pertaining to companies has proliferated through the creation of the World Wide Web. More data means that human analysis is more difficult; but digitized data means that machines can play a role in alleviating the analysis burden. Investment professionals evaluate a wide variety of often contradictory information to decide whether a share is investable or not. There is inherent uncertainty present in share evaluation and investment decision making. Investment professionals will engage in complex analysis and modelling in an attempt to reduce this uncertainty. However, given that stock market data is highly time variant and normally follows a non-linear pattern predicting future share performance can be challenging.

This research investigates the efficacy of using ontologies and Bayesian networks for automating share evaluation on the JSE. Bayesian networks are especially well suited to this research problem given that they are able to represent uncertain and ambiguous knowledge that investment professionals often deal with in their evaluation of shares.

Different perspectives and investment approaches have presented challenges for existing research to effectively represent knowledge pertaining to investment decision making and share evaluation. It is for this reason that this research focuses on a specific investment approach to provide the context necessary to build a concrete and useful model for share evaluation. In addition, a specific investment approach means that the model may be effectively grounded in investment theory and expert practice.
Value investing has been selected as the investment approach with which to model share evaluation. One of the most remarkable regularities in empirical literature has been the fact that value investing is consistently associated with positive abnormal share returns [12]. These studies have proven that firms trading at lower price multiples, stronger balance sheets, more sustainable cash flows, higher profitability, lower volatility, lower Beta and lower distress risk earn higher future stock returns. This pattern of higher future stock returns is referred to collectively as the “value effect” which was first recognized and documented by Benjamin Graham as early as 1934 [1]. Value investing [1] adheres to the “value effect” by buying or selling shares on the basis of a perceived gap between their current market price and their fundamental value.

Existing research which explores knowledge representation with respect to share evaluation requires refinement with respect to adequately reflecting investment theory and expert practice. While the investment decision is complex; at the very least a simplified predictive decision model that reflects some of this complexity should be captured. A predictive model with respect to share evaluation could be used to help investment professionals decide which shares are investable for inclusion in an investment portfolio.

1.1. Aims and Objectives

The aim of this study is to capture and reason with expert knowledge with respect to the share evaluation process to predict future share performance following a value investing approach through the use of ontologies and Bayesian networks. The research aims to capture knowledge on the share evaluation process and more specifically how this would be carried out realistically in a usable application. Hence, share evaluation was explored within the context of a specific investment approach, value investing, to ground the abstract process and concepts to a concrete application.

The objectives of the research are:

- Review existing research studies in the financial domain to analyse the decision process that underpins future share return prediction through a value investing approach.
- Develop an ontology which captures domain knowledge for share evaluation under a value investing approach.
- Develop a Bayesian network to capture the decision process for share evaluation under a value investing approach.
- Evaluate the ontology and Bayesian network for predicting future share performance.
- Develop and evaluate a prototype system for predicting future share returns using a value investing approach.

The expected outcome of the research is:

A system design, hereinafter referred to as the "INVEST system", of which selected key components are implemented through a prototype model. These key components are:

(1) **An ontology** to represent and capture the knowledge of experts who carry out share evaluation under the value investing approach;
(2) **A Bayesian network** that can reason with this knowledge to make decisions pertinent to share evaluation.

The model shall not be exhaustive with respect to share evaluation; the decisions and knowledge in this arena are complex and infinite. Thus, the model has been implemented for the value investing approach. The model should provide a framework for extension for other bottom-up investment approaches.

1.2. Tools and Approach

The research approach requires that knowledge on share evaluation under the value investing approach be synthesized and verified through a review of the literature and consultation with investment experts in professional practice to ensure the basis of the decision model is theoretically sound. A review of current literature that outlines decision models which focus on share evaluation led to the identification of current state of the art proposals which were used to inspire the design of the decision model in this research. Existing implementations of ontologies were re-used where possible in the current model design, such as those contained in the FIBO and SONAR financial ontologies. The ontology was designed using Protégé while the Bayesian network was designed using Netica. An ontology engineering approach, the Unified Process for Ontology Building, was used to guide the design process of the ontology and Bayesian network. The resulting decision model for share evaluation was refined through extensive examination of the literature and consultation with domain experts. The predictive performance of the decision model was validated through back testing using real-world historical financial datasets for shares on the Johannesburg Stock Exchange (JSE) from Bloomberg. Evaluation measures used as good indicators of success include return and risk-adjusted return measures.

1.3. Contributions

The primary contribution of this work is to make explicit the share evaluation process as carried out by a value investor through the creation of an ontology and Bayesian network developed through a methodological framework that is well documented. A cohesive ontology and Bayesian network with respect to share evaluation will allow for use of digitized analysis, automated decision making and less reliance on human processing and endeavour to reach investment decisions. Investment professionals can spend more time digging deeper into understanding the company at hand if a large portion of the analysis is automated. In addition, while this ontology remains grounded in current theory and practice, alleviating the investment professional of tedious analysis may open up further profound and accurate ways of evaluating companies and reaching investment decisions.

1.4. Structure of the Dissertation

A review of related work is presented in the Literature Review in Chapter Two. This is followed by an analysis of experimental design methods in Chapter Three. Chapter Four describes the design and implementation of ontology and decision model using Bayesian networks. This design is validated, the results of which are presented and discussed in Chapter Five. A summary of key findings, a discussion of possible future work and the study's conclusion are presented in Chapter Six.

2. Literature Review

This chapter serves as critical assessment of related work. It begins with an outline of the stock market, modern portfolio theory succeeded by a discussion on portfolio management and share evaluation approaches. An evaluation of general decision support systems follows which serve to outline existing automated approaches to share evaluation and inform the final system design. This research is focused on share evaluation following the value investing approach which is discussed in detail. Further to this, ontologies are proposed for automated share evaluation and thus an outline of ontologies, ontology engineering and ontology engineering methodologies follows. Bayesian networks are introduced so that the limitations inherent to ontologies may be alleviated through their use in the final system design. The chapter ends with a discussion which underscores the need for ontologies and Bayesian networks to be used in combination to create semantic Bayesian networks which may be used to successfully model the share evaluation decision.

2.1. The Stock Market

The stock market enables buyers and sellers to negotiate prices and make trades on listed securities. The stock market works through a network of exchanges, for example, the Johannesburg Stock Exchange or the Nasdaq. Companies list shares of their stock on an exchange through a process called an initial public offering ('IPO') which allows the company to raise money to grow the business. Following this, investors may trade shares among themselves. The price as tracked on the exchange is guided by supply and demand of each listed stock. The goal of an active investor is to exploit mispricing in stock prices which arise as a result of stock market inefficiencies. Market inefficiencies are said to arise when all publicly available information regarding the stock is not fully reflected in the price, meaning that the price is not reflective of the stock's true value.

2.1.1. Modern Portfolio Theory: The Efficient Market Hypothesis

The Efficient Market Hypothesis ('EMH'), as first proposed in the 1970s [2], is an investment theory whereby share prices reflect all available, relevant information in a rapid and unbiased fashion [3] [4] and are reflective of a share's true value. The EMH means that market inefficiencies and thus the ability to earn abnormally high risk-adjusted returns through active investing is impossible. The EMH finds its origins in the Random Walk Theory which prescribes that new information is randomly favourable or unfavourable relative to expectations intimating changes in share prices in an efficient market should be random [5].

However, evidence exists that proves the stock market is not as perfectly efficient as the EMH suggests. It has been proven empirically that shares with low price-to-earnings ratios, on average, earned higher absolute and risk-adjusted rates of return than shares with high price-to-earnings ratios [4]. This finding proves that information is not always 'fully reflected' in security prices in as rapid a manner as the EMH suggests. Fama [3] synthesized existing research by defining three different forms of market efficiency: weak form, semi-strong form and strong form. The three forms of market efficiency build on each other and are defined in terms of the rapidity and accuracy of a price adjustment to new information. The three forms are as follows: (1) weak form assumes no dependence in stock price movements meaning that historical information on price movements will not impact future price movements; eliminating profit opportunities through share price trend analysis (known as 'technical analysis') while (2) semi-strong form posits that all publicly available and relevant information is reflected in the stock price;

further eliminating fundamental analysis and (3) strong form states that all privileged information is also reflected [3]. Grossman and Stiglitz [6] recognized that the premise on which an extremely high level of market efficiency rests is internally inconsistent. Profit opportunities necessary to incentivize security analysis would be eliminated. Security analysis and subsequent trading in mispriced shares make the market efficient as trading moves the prices towards the stock's true value [5]. 'Market frictions' like the cost of security analysis and trading, limit such market efficiency. Market efficiency should be seen on a continuum as opposed to the exact forms; "fully efficient" is an exacting standard that is unlikely to be met in the real world [6]. Where greater information efficiency exists, a market is likely to be more efficient.

There is abundant literature on persistent market inefficiencies in the South African market which are termed "stylized facts" in the literature [6 – 8]. These persistent inefficiencies incentivize active investors to perform share evaluation to earn abnormal risk-adjusted returns relative to the market over the long run. This activity is discussed in further detail below.

2.1.2. Portfolio Management: Share Evaluation

Share evaluation is the initial decision-making phase of the portfolio management process. Portfolio management is concerned with building an efficient portfolio through the correct investment choices and is one of the most important problems encountered in the field of financial engineering. Portfolio management is comprised of two procedures: **asset evaluation**, to identify assets that are expected to generate an attractive return and **portfolio construction**, to select the assets and determine the percentage in which each asset shall participate in the portfolio. The investor aims to choose a set of assets for a portfolio that maximize return within a chosen risk profile. Asset evaluation and selection consists of identifying markets, market sectors and individual assets that will do better or worse than the rest, and over- and underweighting them in portfolios [9]. Share evaluation is used to identify individual shares that have desirable risk-return characteristics such that a list of attractive shares is available for potential selection in the portfolio construction phase [9]. The portfolio construction phase is concerned with the manner in which an optimum portfolio is constructed through the correct combination of attractive securities as identified in the security evaluation phase that also fit together based on their correlations. The aim of portfolio construction is to hold sufficient independent positions in securities to avoid the concentration of risk. Shares are often grouped into their homogenous sectors in order to assess sectoral risk concentration. Portfolio construction is beyond the scope of this research but it has been briefly mentioned to place share evaluation in context of the larger portfolio management problem.

2.1.3. Share Evaluation Approaches

Share evaluation is employed by active investors. There are several investing and share evaluation approaches; common investment approaches are discussed below.

The **technical approach** focuses on plotting the price movement of a share and drawing inferences from this price movement. The basic assumption of all technical theories is that history tends to repeat itself; past patterns of price behaviour in individual securities will tend to recur in the future. This assumption is predicated on the idea that successive price changes in individual securities are dependent. Keynes [10] presented research which supports the use of the technical approach; noting stock prices are based on speculation rather than economic fundamentals and stock markets are little more than casinos with no economic role. While this view may hold in the short run, prices driven by speculation converge in the long run to those that would exist based on fundamentals. Unlike technical approaches which may not hold in the long run, fundamental approaches hold as markets are not entirely efficient.

Fundamental approaches evaluate various factors pertaining to shares beyond price such as profitability, quality of management and growth [11]. Two common **fundamental approaches** are **growth investing** and **value investing**. The distinction between the two approaches is largely based on the investment premise. **Value investing** is premised on the idea that undervalued shares will deliver an investment return greater than the market return. Analysis of these fundamental factors should lead to the determination of an intrinsic value which is based on the expected earning potential of a share and may be above or below the current share price. On the other hand, **growth investing** is premised on the idea that companies with exceptional growth prospects will deliver a return greater than the market return and reflect a preference for companies with strong earnings growth, research and development intensity, and innovative technology. In contrast to value shares, growth shares will usually have a high price-to-earnings ratio, a high price-to-book ratio and generally do not appear to be undervalued. These investment strategies are not mutually exclusive given that the premise of the one does not eclipse the premise of the other and some investment professionals employ both strategies.

This research focuses on **value investing** given that it is the popular and dominant approach in the investment domain. Before delving further into this in Section 2.3, the section below explores decision support systems that currently serve to automate share evaluation.

2.2. Decision Support Systems for Share Evaluation

Most decision support systems (DSS) for share evaluation are unsurprisingly multi-criteria in nature given that they serve to tackle a multi-criteria decision problem which requires numerous factors to be evaluated to reach a final decision. Multi-criteria decision support systems (MCDSS) for share evaluation do not explicitly focus on a specific investment approach but are generally modelled on the fundamental analysis technique which is carried out by many investors in practice.

2.2.1. Review of Multi-criteria Decision Support Systems

Aouni et al. [28] performed a bibliographic review of the current work to date on MCDSS for portfolio selection and more specifically, for share evaluation utilizing exact methods. A summary of this review has been detailed below and shall inform the design of a more robust system which shall utilize ontologies and Bayesian networks to model the share evaluation process. Exact methods guarantee that a provably optimal as opposed to approximate solution to the share selection problem is found. Their review was carried out to highlight the increasing importance of multiple criteria decision aid methods for portfolio selection; and reflect a large category of decision support systems (DSSs) that have been created utilizing a multi-criteria decision-making framework.

The review highlights existing works are focused on identifying evaluation criteria and determining their relative weights for the share selection process. These works attempt to capture the nature of the share evaluation process as carried out in practice which has expanded beyond the standard return and risk factors to incorporate other fundamental factors that investment professionals must weigh up.

Further to this, the review highlights that the criteria considered in share evaluation and selection have increased markedly in number and variety over recent years. The necessity has arisen to enhance existing decision-making models with new approaches focused on individualism and realism [28]. Upon inspection of the works reviewed by Aouni et al. [28], the multi-criteria decision models lack the realism necessary to be employed in practice. Many of the works fail to reflect the sequential nature of share evaluation and ignore important share evaluation factors and concepts like intrinsic value. This stems from a lack of thorough scientific research and consultation with investment professionals on share evaluation. The use of disparate share evaluation factors or criteria across MCDSS and no clear categorization of these criteria reiterate the need for an ontology to serve as a clarifying artefact for both financial and non-financial users. Research in the area of multi-criteria decision-making pertaining to portfolio management continues to be active and increasing and is poised to take on a variety of emerging issues such as the consideration of new risk measures, the role of non-financial dimensions (such as sustainability and social responsibility) and large-scale datasets from global markets. Aouni et al. [28] point out that while additional criteria provide added value; they also present an enriched set of criteria which require complex trade-offs. Sometimes the inclusion of additional criteria and constraints can make the share evaluation problem too complex to be solved by existing exact methods [28]. While multi-objective evolutionary algorithms are an alternative to these exact methods; additional algorithmic risk and an approximately accurate solution is often untenable for asset managers with billions of dollars under their stewardship. These reasons highlight that exact methods remain necessary but that the criteria included in the model must be scrutinized so as not to overburden the decision process.

Several multi-criteria decision aids (MCDAs) have been used by researchers to evaluate the criteria and rank securities accordingly; the most popular being the analytic hierarchy process (AHP), followed by ELECTRE-based approaches and TOPSIS approaches [28]. These approaches are used to rank criteria according to their importance as believed by decision makers. AHP, for example, is interested in the pairwise comparisons between criteria using a scale from 1 to 9 and aggregating these relative priorities in order to rank alternative options like shares. Many MCDA like AHP can and have been used to provide objectified inputs for the Bayesian network given that it allows decision makers to objectify their decisions and formalize the decision process through pairwise comparison of criteria. While MCDA like AHP may serve to complement and aid in the design of the Bayesian network, Bayesian networks remain superior for making predictions about uncertain factors like future share price performance [29].

Table 1 below summarizes the criteria used in the reviewed works and illustrates their frequency of use in the studies reviewed by Aouni et al. [28]. As illustrated below there are a broad set of criteria which may be used as share evaluation factors in decision models but it appears unclear as to which are the most important per existing research or how these should be organised. The variation in criteria pose a challenge for representing common and sharable share evaluation factors in decision models. It appears that the variation in factors stem from the lack of clear objective present in existing decision models in research. Existing decision models employ too broad an objective, simply stating ‘share evaluation’ as the overarching objective which makes it difficult for the modeler to hone in on the most important and relevant share evaluation factors. Honing in on a specific investment approach like value-investing, growth or momentum means that the relevant factor set may be identified as each investment approach focuses on specific factors or weights them differently in the decision process. It is key to be clear as to the investment approach under which the share evaluation process is being conducted to have a clear decision model and framework. This research aims to address the issue of poor deliberation over share evaluation factors through development of a conceptual model with a new categorization framework.

This will ensure that the decision model is not overburdened with ineffectual share evaluation factors and provides the modeler with a clear means by which to select relevant evaluation criteria for inclusion in the decision model. Further to this, the model design in this research focuses on the value investing approach to illustrate the selection process of evaluation factors as carried out in practice where the objective of the decision is specific.

Table 1: Share Evaluation Factors and their use in various MCDA studies

Share Evaluation Factors	Number of Studies
Expected return (mean historical returns)	13
Other criteria (industry evaluation macroeconomic factors, efficiency of management, investors preferences)	9
Environmental, social and governance (ESG) aspects	6
Return on equity (ROE)	5
Risk (variance, standard deviation)	5
Risk (Beta)	4
Total assets turnover ratio	4
Price/earnings ratio	4
Equity/debt ratio	4
Market share	3
Return on assets (ROA)	2
Lower semi-variance	2
Long-term debt to equity ratio	2
Total debts to total assets ratio	2
Times interest earned	2
Inventory turnover ratio	2
Profit margin	2
Market value	2
EPS coverage	2
Liquidity	2
Gross book value per share	2
Capitalisation ratio	2
Marketeability	2
Financial position progress	2
Exchange flow ratio	2
Round lots per day	2
Transaction value per day	2
Equity ratio	2
Structure ratio	2
P/B ratio	2
Loans to deposit ratio	2
Net Asset Value to Enterprise Value ratio	2
Dividend yield	1
Expense ratio	1
Manager’s tenure	1
Sharpe ratio	1
The maximum rate of acquisition or liquidation costs compounded by the fund	1
The fund’s eligibility for the real effective exchange rate	1
Transaction costs	1
Skewness	1
Kurtosis	1

Share Evaluation Factors	Number of Studies
Result of audit	1
Production growth	1
Sales growth	1
Payout ratio	1
Profitability variance	1
Free float of the stock	1
Capital raising within the last year	1
Price to fair value ratio	1
Price per share	1
Price to cash flow ratio	1
Price to adjusted after-tax operating earnings ratio	1
Economic value added	1

2.2.2. A MCDSS for Share Evaluation Using Fundamental Analysis

Post review of the papers highlighted by Aouni et al. [28], the MCDSS presented by Samaras et al. [30] appeared to be the most relevant to the research problem. The work presents a decent proxy for how most MCDSS have been formulated. The objective of the work by Samaras et al. [30] was to present a real-world system to be employed by institutional and private investors that could evaluate shares on the Athens Stock Exchange (ASE), on the basis of fundamental analysis. The system utilizes quantitative and qualitative data and multi-criteria analysis methodologies in order to rank shares from best to worst to single out ones eligible to be included in the portfolio being constructed. The evaluation considers the specific characteristics of the potential investor and is able to incorporate investor choice or preference; the investor is able to select their level of undertaken risk which is an investment choice rather than a simple evaluation criterion.

Stock ranking based on fundamental factors is carried out on the basis of distinction between four accounting plans; often referred to as *contextual fundamental analysis* which allows one to compare firms from the same industry. The four accounting plans are Commerce/Industry, Banking, Insurance and Investment. Within each accounting plan evaluation there is a different evaluation structure with different criteria sets; taking into account the nuances between the different sectors. Samaras et al. [30] dive into identifying industry specific evaluation factors but fail to identify universal evaluation factors common to all shares. This research takes a step back to identify these universal evaluation factors, which could then be extended to include industry specific factors.

The model presented by Samaras et al. [30] is an important one as it illustrates the incorporation of investor preference and variable thresholds in modelling through the use of the Lowest Stock Profile (LSP) against which values for fundamental factors may be evaluated. The LSP describes the lowest standards a stock should meet in order to be considered worth investing and to be included in the stock portfolio [30]. All shares that are below the LSP are cut off from the list of "acceptable shares" (i.e. those eligible to be included in the in the portfolio to be formed). There are three levels of LSP, which have been determined to encompass different investor preferences: *regular* for the balanced-indifferent investor, *strict* for a conservative-demanding investor and *lenient* for a dangerous-slack investor. Each has different cut-off points for "acceptable" investment shares. These three levels of the *LSP* constitute reference sets that are inserted into the multi-criteria ranking model which employs UTASTAR to then categorize the investment shares based on the *LSPs*. The UTASTAR model contains the marginal utility function which has been determined from the expert decision maker's ranking of companies from best to worst based on the criteria presented and his preference profile. Thus, the output of this procedure will be the proposed ranking for the model that expresses the way that the model decodes the criteria. Samaras et al. [30] provide an example of how to create a flexible decision model which may be adjusted to suit the specific investor; these features shall be incorporated into the model design in this research.

Samaras et al. [30] highlight the necessary interactive minimization effort between an expert and the model to reduce over- and under-estimation errors between the ranking of the decision maker and ranking of the model. A few iterations of the minimization effort lead to a satisfying convergence between the model ranking and decision ranking. Following this minimization effort, the estimated model has essentially been approved and can be extended to evaluate further sets of companies. A similar method of evaluation and refinement is carried out in this research.

As a final point; neither Samaras et al. [30] nor the works highlighted by Aouni et al. [28] make explicit the model design or knowledge acquisition process. It is clear that the model design and share evaluation factors included in this design should be extricated from investment literature and what is carried out in investment practice for the selected investment approach. The design approach in this paper is more rigorous than in existing research; an extensive literature review of value investing theory and practice has been performed to inform the conceptual model. This research is presented in the section below.

2.3. The Value Investing Approach for Share Evaluation

Share evaluation can take many forms but most investors will conform to a specific style of investing to guide their analysis and decisions. This research focuses on value investing, an investment style first conceived by Graham and Dodd and a vital discipline employed by most active managers. A simplistic definition of the value investing approach refers to the buying or selling of shares on the basis of a perceived gap between their current market price and their intrinsic value [12]. Intrinsic value is commonly defined as the present value of the expected future payoffs to shareholders [5] and thus is representative of a share’s long-run value. The value investing approach rests on the fact that prices of financial securities are subject to unpredictable swings due to over-reaction by stock market participants to good or bad news but, despite these swings in prices, most securities have underlying intrinsic values that are relatively stable and can be measured with reasonable accuracy by a disciplined and diligent investor [13]. According to the simplistic view of the value investing approach, an investor who buys securities only when the market price is significantly below the calculated intrinsic value will earn superior returns in the long run. The perceived gap between the market price and the intrinsic value is known as the margin of safety and is employed by investors to account for the inherent uncertainty in estimating the intrinsic value. Given that intrinsic value is based on multiple assumptions and unknown facts, it can never be determined exactly and thus there is always room for disagreement among market participants as to what the intrinsic value should be [12]; hence there will always be a buyer and a seller of the same share.

An enormous body of literature documents the success of the value investing approach, which is the tendency of value shares (shares with low prices relative to their fundamentals) to outperform glamour shares (shares with high prices relative to their fundamentals) [12]. While academics most often agree on the empirical facts, there is much less consensus on the reason behind these findings [12], but the broad tendency of value shares to outperform glamour shares is quite a robust finding in academic literature. Some believe value shares are indicative of a bargain; others believe value shares are cheap for a reason and that common measures of value are also indicators of some sort of risk. It is for this reason that most value investors look beyond perceived “cheapness” to evaluate further fundamental factors prior to making a stock specific investment decision. Given this, value investing is often portrayed too simplistically and is in fact a multi-attribute problem extending beyond intrinsic value evaluation.

Lee [12] articulates that value investors are concerned with two primary categories of share evaluation factors, namely “cheapness” and “quality”. His work discusses the Levin-Graham score [14] to provide empirical evidence that stock screening based on factors pertaining to value and quality continue to accurately predict stock returns [12]. Value is often referred to as “cheapness” and this is perhaps too restrictive to the work carried out by value investors who prefer to (1) find quality companies and (2) buy them at “reasonable prices” [12].

Simply put, **Value Investing = Reasonable Price + Quality**

This overarching theme of value and quality is extremely helpful when trying to understand the investment approaches of famous investors’ like Warren Buffet, Charlie Munger and Joel Greenblatt. A recurrent theme of Buffet’s investment style was to look for quality companies at reasonable prices. Joel Greenblatt studied and synthesized Buffet’s strategy of investing in the *Magic Formula* [15]. This remarkably simple strategy ranks companies on the basis of just two factors: return on capital (ROC) and earnings yield (EY). The formula simply states one should look for companies with ROC of 20% or higher over the past 5 years (the “quality” factor) and from these, pick those ones with the highest earnings yields (EY) (the “value” factor). This formula has been proven to work with firms ranked top by the formula outperforming their peers by a wide margin over the past 50 years. This is not very different to Benjamin Graham’s screening tool formulated many years earlier which focused on five years of high and consistent growth (the “quality” factor) and low P/E ratios (the “value” factor). The verdict from investors in the field: value combined with quality pays.

As previously highlighted share evaluation is a multi-attribute problem. Fundamental analysis is the technique used by value investors to deal with this multi-attribute problem and refers to the use of historical accounting numbers to determine the intrinsic value of a stock and other factors for prediction of returns [12]. The technique is ‘bottom-up’ which effectively means that an investor will evaluate the company’s financial situation and outlook, regardless of the general economy or sector. A number of recent studies provide compelling evidence that historical accounting numbers are informative and are already playing a useful role in fundamental analysis and investing [12].

2.3.1. Share Evaluation Factors

Previous studies indicate that the stock evaluation process is highly unstructured and there is not a fixed set of factors for evaluating shares [16]. There are many factors, perhaps an infinite amount, which an investor may choose to evaluate a share. That being said, it is imperative to understand that the investment approach guides the selection of factors for the evaluation of shares. Under the value investing approach, it is generally understood that two primary categories of share evaluation factors are used by value investors, namely “value” and “quality” [12]. Context of the investment approach, which determines the investor’s objective, provides the framework for factor selection.

Examination of existing research provides a deeper understanding of the share evaluation factors employed for value investing using fundamental analysis. Stock evaluation under the value investing approach can be articulated successfully through combination of investment theory, on the one hand, and experience of long-time field professionals, on the other. This combination has been detailed below.

2.3.2. Value Factors

It must be understood that value shares are not just those that are cheap relative to current capital in place (book value) but include those that are cheap relative to the present value of their future residual income [12] or some forward-looking cash flow or earnings metric; where future growth prospects are accounted for. Cheapness of a share is confirmed when the intrinsic value of a share is below the current share price by a reasonable margin of safety. The intrinsic value of a share is the total estimated value of equity divided by the number of shares and is a value that cannot be improved to yield a better value [17]. Valuation models are used to determine the intrinsic value of a share.

Two basic models are used to estimate intrinsic values: dividend discount models (DDMs) and free cash flow approaches (FCF). The dividend discount model focuses on dividends while the free cash flow model focuses on sales, costs, and free cash flows [18]. An alternative approach is the use of models based on multiples like the price-to-earnings ratio [18].

There is a pervasive scepticism by practitioners regarding formal valuation models like FCF or DDM, with many often discarding them in favour of 'rough-cut' methods such as pricing on the basis of comparable firms and simple price-to-earnings ratios [17] [18]. Penman [17] points out that valuation as taught in theory often gives the impression of precision. The free cash flow model and dividend discount models are significantly affected by the estimated growth rate of the variable used (dividends, earnings, cash flows or sales) and are affected by the investors required rate of return on the stock. Both of these variables must be estimated. Imprecision can arise from small changes to inputs like the discount rate or growth rate can have a large effect on the assessed valuation. Graham and Dodd [13] recognised that while we are concerned with the discrepancy between intrinsic value and price; intrinsic value remains an elusive concept. While intrinsic value is felt to be a value as justified by the facts, it would be remiss to imagine that intrinsic value is as definite and determinable as the market price.

While several approaches and associated models exist for valuation in theory; we explore intrinsic valuation as carried out in practice below. While financial theory provides tools and models, the implementation of valuation techniques allows for considerable latitude to professionals [19].

Pinto et al. [20] explored equity valuation in professional practice through their survey of a subset of Chartered Financial Analysts focused on equity valuation in professional practice. This subset worked across brokerage firms, hedge funds, investment banks and investment management firms; all of which were well-presented. The survey was a global effort with a respondent sample split between the Americas (66%), Asia Pacific (12%) and Europe, Middle East and Africa (22%).

Pinto et al. [20] gathered information on the conditional frequency of use of valuation approaches; that is, the percentage of cases in which a given approach was reported to be used, given that the analyst previously indicated that he or she uses the approach. Their survey rendered results which displayed an overwhelming preference for market multiples valuation; much in line with Penman's observations on 'rough-cut' approaches. There is also widespread adoption for the present discounted value approach and asset-based approaches. In practice, there is limited adoption of real option approaches (12.7% of respondents) and other approaches (5.0% of respondents). The pertinent results of the survey are displayed below in Table 2 below which identifies the most common valuation approaches and the specific valuation models employed in professional practice. It is important to note that these valuation models are used to estimate the intrinsic value, either in relative terms through market multiple approaches or in absolute terms through discounted value approaches. The issue with some models is that they only compare the price of a stock to its capital in place (book value) but fail to account for growth opportunities. Price-to-book, price-to-cash flow and price-to-sales focus on current capital in place as opposed to more forward-looking multiples which account for these growth opportunities. It should not be construed that these multiples are not relevant but rather that these remain supplementary to an investor's analysis.

Table 2: Survey of most common Valuation Approaches and Models employed in professional practice

Valuation Approaches	**Percent of Respondents**	**Percentage of Cases Respondents Use Each of the Approaches**	**Valuation Model/s**	**Percent of Respondents**	**Percentage of Cases Respondents Use Each of the Approaches**
Market Multiples Approach	92.8	68.6	Price-to-Earnings (some measure of earnings)	88.1	67.2
			Enterprise Value	76.7	61.1
			Price-to-Book	59.0	44.8
			Price-to-Cash-Flow (some measure of cash flow)	57.2	54.6
			Price-to-Sales	40.3	45.7
			Dividend Yield	35.5	44.3
			Other Ratios e.g. PEG ratio	11.6	58.5
Present Discounted Value Approach	78.8	59.5	Discounted Free Cash Flow Model	86.8	80.1
			Dividend Discount Model	35.1	51.7
			Residual Income Approach (e.g. based on discounted abnormal earnings, economic profit, EVA, or similar concepts)	20.5	46.1
			Cash Flow Return on Investment (CFROI)	19.7	58.5
			Other Model	3.6	71.3
Asset-based Approach	61.4	58.1	Book Value, Adjusted Book Value, Asset Market Values or Asset Replacement Costs	Not disclosed in the study	Not disclosed in the study

The most widely used and generally applicable valuation models observed in practice are the price-to-earnings multiple and discounted cash flow model [20] which are discussed below. While the enterprise value multiple is used by a large percentage of respondents, the enterprise value multiple is not well suited for valuation of financial services firms [97] and was not selected.

2.3.2.1. A Market Multiple Approach to Intrinsic Value Determination

Price-to-Earnings ('PE') Multiple

The price-to-earnings multiple ranked the highest across all regions of Pinto et al. [20] and hence has been considered for discussion. The PE multiple indicates the rand amount an investor can expect to invest in a company in order to receive one rand of that company's earnings. PE is a universally meaningful factor for evaluation of companies. On the other hand, valuation multiples like the enterprise-value-to-EBITDA, price-to-book, price-to-cash flow and price-to-sales cannot be meaningfully applied to companies from certain sectors and thus are difficult to easily use in the decision model. In contrast, PE is powerful in that it allows comparison to the market, sectors, peer firms and a share's own history.

Table 3: Inputs for PE Multiple

Valuation Approach	Valuation Model	Possible Inputs	Preferred Inputs by Respondents
Market Multiples Approach	Price-to-Earnings ('P/E')	**Price** (Numerator) **Earnings** (Denominator) • Trailing net income • Trailing operating income • Forecasted net income • Forecasted operating income	**Price** **Earnings** • Forecasted net income

The PE ratio was first presented by Benjamin Graham and David L. Dodd in 1934, [20] survey and several other surveys confirm its continuing primacy in practice. In effect, 81.8% of the respondents in [20] survey use the PE ratio in the course of their valuation work. It must also be highlighted that forward-looking PE ratios were preferred among respondents to trailing PE ratios, with net income preferred to operating income in measuring earnings. The dominant definition of PE, as preferred by 61.1% of respondents using a PE multiple, was a forward P/E based on forecasted net income. Analysts appear to follow theory in that forward-looking earnings are, all else equal, more relevant than past earnings. A forward PE indicates what the market is willing to pay for a stock based on its future earnings. A high PE could mean that a stock's price is high relative to future earnings and is possibly overvalued. Conversely, a low PE might indicate that the current stock price is low relative to future earnings.

PE Relative

A further derivative of the PE ratio is introduced to ensure that it remains meaningful when employed in fundamental analysis. A common error in the utilization of PE for a share is to only compare it to the mean PE for the market, for example the JSE All Share Index, and deduce whether the price is reasonable or cheap relative to the market. The pitfall of using PE in this manner is that it assumes the share's PE mean reverts to the mean PE for the market over time. It does not consider how the share trades relative to the sector or the market over time; it is known that certain shares always trade above or below the mean of the market or the sector due to their company or sector specific characteristics. For this reason, a PE relative ratio should be used as an input for the evaluation of whether a share's price is reasonable relative to the market or the sector. The use of relative fundamental factors was illustrated in [16] where the general rule for calculation of relative ratios is as follows:

$$\textit{Relative fundamental factor} = \frac{\textit{Fundamental factor}}{\textit{Corresponding industry or market average}}$$

Relative fundamental factors allow for shares to be evaluated and compared regardless of their sector classification. Thus, the PE relative to the market or the sector should be compared to its own history over time. A PE is only applicable when comparing a share's current or forward PE to its historical PE given that an investor would expect a share's current or forward PE to mean revert to its historical PE.

Universal Share Evaluation Factors

In their research paper Yunusoglu [16] outline their use of share evaluation factors which are universal to all shares. Their work highlighted that even though one may be able to identify universal share evaluation factors for shares across different industries, the mean values for these factors may differ across industry classes due to different characteristics of the industries. Yunusoglu's [16] proposes the use of relative fundamental factors to alleviate this issue. Relative fundamental ratios alleviate the issue presented by comparing shares from different industries which may have different mean values for fundamental factors which do not converge to a universal mean. The PE relative ratio serves as an example given that a stock's PE relative to a sector or market PE may never converge towards the same mean and controls for the fact that the share price may have increased due to market or sector movements rather than stock specific factors.

Where appropriate, this work employs Yunusoglu's [16] proposal of using relative fundamental ratios to evaluate shares. While Yunusoglu's [16] approach is to convert all ratios or evaluation factors to relative ratios, this work has not done the same. Some factors should be evaluated in absolute terms. ROE serves as an example of why this is the case. An investment professional will generally demand a minimum ROE and at the very least one that exceeds the COE. A relative ROE may disguise the fact that the ROE on an absolute basis is poor. This work shall build on the approach proposed by Yunusoglu [16] to illustrate a more nuanced approach to how evaluation factors are used; either as absolute or relative values depending on the factor in question.

2.3.2.2. A Present Discounted Value Approach to Intrinsic Value Determination

Present discounted value approaches follow from the fact that the value of an investment is presented by the future consumption that it is expected to yield, discounted for both the time value of money (i.e. the risk-free rate) and the risk that the expectation may not be realised. Given that cash buys this consumption, the value of an equity investment is given by the expected net cash dividends, with a discount applied. This is expressed via the Dividend Discount Model. While the correctness of this model is widely accepted, issues confront this valuation model. See [21][17] for a discussion of the dividend discount model limitations like the dividend irrelevance proposition which states that the intrinsic value is insensitive to the payment of dividends up to any point prior to a liquidating dividend, given that dividends represent value distribution rather than value generation. The Discounted Cash Flow (DCF) valuation deals with some of the issues articulated by substituting the dividends for measures that pertain to value generation, namely free cash flow.

Discounted Cash Flow ('DCF') Model

The most widely used Present Discounted Value Approach across all the regions is the Discounted Free Cash Flow model with other discount value models far less frequently used [20].

Table 4: Inputs for DCF Model

Valuation Approach	Valuation Model	Inputs
Present Discounted Value Approach	Discounted Cash Flow Model	**Cash Flow** **Discount Rate** • Required Return on Equity • Equity Risk Premium • Risk free rate **Terminal Value**

There are several choices that must be made by an analyst when employing a DCF model. The first choice is concerned with the approach for estimating the required return on equity (i.e. the cost of equity); [20] indicates respondent's choices in their survey which has been tabled below.

Table 5: Preferred approaches for determination of required return on equity

Approach for estimating the required return on equity?	Percent of Respondents	Percentage of Cases Respondents Use Each of the Approaches
The Capital Asset Pricing Model (CAPM)	68.2	77.5
A judgementally determined hurdle	47.5	64.3
Bond Yield plus a Risk Premium	42.7	61.4
Arbitrage Pricing Theory	4.8	47.0
Fama-French or related model	4.0	42.2
Other	6.3	79.4

The second choice pertains to how to an investor determines an equity risk premium. Pinto et al. [20] indicates the estimates used to determine the equity risk premium which may be part of a required rate of return computation. Bancel and Mittoo [19] established that while the valuation method may be agreed, no best practice exists for the estimation of parameters used within an approach like the CAPM.

Table 6: Preferred approaches for estimating an Equity Risk Premium

Approach for estimating an Equity Risk Premium	Percent of Respondents	Percentage of Cases Respondents Use Each of the Approaches
Historical equity risk premium	36.2	77.7
Forward Looking Equity Risk Premium	34.7	75.3
Adjusted historical equity risk premium	26.9	72.6
None. I do not estimate an Equity Risk Premium	24.5	N/A
Other	4.8	82.4

It becomes evident that the DCF is computationally burdensome given the numerous inputs and assumptions required. Further to this, given that intrinsic value renders an absolute value it provides no means by which to compare shares from a similar industry or sector easily.

2.3.3. Quality Factors

The rhetoric of most value investors is not only to focus on ‘cheap’ shares but rather to find “quality” shares (firms with strong fundamentals) trading at reasonable prices. The academic evidence in favour of quality investing is less explicitly studied and difficult to recognise. There is no clear view agreed by academics on what quality might look like. For example, many research papers have examined the persistence of earnings or the ability of accounting-based variables to predict future returns, but most have not done so explicitly under the quality rubric [12]. That does not mean it is not proven to be effective in predicting higher returns. Novy-Marx [22] illustrated that profitable firms generated significantly higher returns than unprofitable firms despite having significantly higher valuation ratios. One of the first conceptions of an intuitive recognition that firms with quality factors like lower leverage, higher liquidity and a high rate of steady growth have a better chance of generating high rates of return in the future was illustrated in Graham’s stock screening tool as presented in *Security Analysis* [13]. This has been followed by many studies which prove firms trading at lower pricing multiples with quality factors like stronger balance sheets, more sustainable cash flows, higher profitability, lower volatility, lower Beta and lower distress risk actually earn higher future share returns [12].

Quality firms are those with a high present value of future residual income or cash flows. Lee [12] looks to determine useful measures for achieving this in future, namely, future profitability and growth which influence future ROE, good cash flows and higher pay-outs while maintaining profitability and growth rates and safety through lower risk and stable earnings. Asness, Frazzini, and Pedersen [23] have pulled together many of the disparate strands for quality investing. In their study, they define quality firms as shares that are “safe, profitable, growing and well managed”. They argue that all else being equal, investors should be willing to pay more for these quality shares but illustrate that the market does not pay a high enough premium for these quality shares. Utilising their quality metric, they create a “quality minus junk” (QMJ) portfolio and find that this portfolio earns positive risk-adjusted returns in 22 out of 23 countries [23]. A summary of quality factors identified in academic research is detailed below in Table 7.

Table 7: Summary of Quality Categories and Evaluation Factors presented in academic research

Category	**Quality Evaluation Factor**	**Academic Research**
Profitability	Return on Assets (ROA)	[23]
	Return on Equity (ROE)	[23]
	Gross Profit over Assets (GPOA)	[23]
	Gross Margin (GMAR)	[22], [23]
	CFOA	[23], [24], [25]
	ACC	[23]
Growth	ΔGPOA	[23]
Financial Risk	Probability of earnings manipulation (M-Score)	[26], [27]

Category	Quality Evaluation Factor	Academic Research
	IVOL	[23]
	EVOL	[23]
	Beta	[23]
	Leverage	[23]
	Financial distress or bankruptcy risk	[23]
Payout	Net equity issuance	[23]
	Net debt issuance	[23]
	Net payout ratio	[23]

As illustrated above, the quality of a stock cannot be determined through one specific metric or ratio; it requires a multi-criteria evaluation given that the performance of a firm is affected by a variety of factors. All else being equal, a high-quality firm is more likely to deliver the expected positive excess risk-adjusted return as opposed to a low-quality firm [12]. Many of the factors detailed in the in the discussion above are recognizable in the review provided by Aouni et al. [28] on multi-criteria decision making for share evaluation detailed in Section 2.2. and as summarised in Table 1. The next sections delve into proposed techniques to represent the knowledge on share evaluation under the value investing approach discussed above.

2.4. Ontologies

While the research and models discussed in Section 2.2. serve to inform the model design in this research; the proposed techniques for representing the knowledge are discussed in further detail below. The discussion which follows provides an overview of ontologies, a Semantic Web technique used to represent and reason with expert knowledge. Ontologies are to be employed in this research to represent and reason with expert knowledge pertaining to share evaluation under the value investing approach.

2.4.1. The Semantic Web

The Web's content was originally designed for humans to read; a medium to house documents and files, but remained difficult for computer programs to manipulate meaningfully. The shift from documents and files to a web of information and data that can be processed automatically and meaningfully signified the extension of the World Wide Web to what has been termed the 'Semantic Web'. The Semantic Web relies on existing technologies, namely eXtensible Markup Language (XML) and Resource Description Framework (RDF) to create this web of data. For the Semantic Web to function, computers must have access to structured collections of information and sets of inference rules, more formally known as *ontologies* in order to conduct automated reasoning. Technologies like XML and RDF support modeling conceptualisations and creation of ontologies for a specific domain of interest on the Semantic Web. Ontologies allow web resources to be semantically enriched; a pre-condition for services like semantic search and retrieval of web resources [31].

To reiterate, the foremost purpose of the Semantic Web is to encapsulate knowledge and its representation so that machines can understand the semantics (i.e. meaning) of information and respond to complex human requests [32] [33].

2.4.2. What is an Ontology?

Ontologies are a modelling tool used to encapsulate and convert knowledge into a machine understandable form [34]. In their simplest form, ontologies serve as a necessary 'common point of reference' for which there is 'one entry per meaningful concept', an accompanying definition as agreed upon by domain experts and an unlimited number of synonyms for each entry [35].

Research probes whether there is a true shared understanding of ontologies given the many definitions that are already in use [33][36][37]. The term is borrowed from philosophy, an ontology being a systematic account of existence. The definition by Gruber [38] is by far the most cited: "An ontology is an explicit specification of a conceptualization". Studer et al. [93] presented a merged definition which removes any need to contend with Gruber [38] and Borst's [39] definition: "An ontology is a formal, explicit specification of a shared conceptualization of a domain". The latter definition focuses on the application of ontologies with respect to computational systems, where **"explicit"** refers to clearly specifying concepts, relations, instances and axioms in the domain of discourse; **"formal"** refers to being machine readable; **"shared"** implies that a community of experts has consented to the content of the ontology; and **"conceptualisation"** implies that it is an abstract, simplified model of a domain; a "set of informal rules that constrain the structure of a piece of reality [36]". Accepted definitions of **"conceptualisation"** emphasize that the conceptualisation should present the consensus rather than the individual view; a shared understanding. Practically, an ontology is typically conceived as a set of terms or concepts (e.g. entities, attributes, processes), their definitions, their connection to each other and some simple rule of inference and logic used within a particular knowledge domain [40][41][42][43]. These components are combined to form a semantic structure which may be used to represent knowledge pertaining to a particular domain of interest.

2.4.3. Types of Ontologies

Within literature, ontologies can be viewed from numerous vantage points. It is not surprising that we find diverse classifications of ontologies [44]. It is imperative to differentiate between various types of ontologies in order to provide clarity regarding their content, their use and their goal. While several classification dimensions for ontologies have been proposed they can be grouped around two broad dimensions: **formality** and **generality** [44]. The decisions pertaining to the formality and generality are often driven by the intended **purpose** of the ontology which should always be considered in ontology classification decisions. Each classification dimension is discussed in turn below.

2.4.3.1. Purpose

At a high-level most ontologies are intended for reuse. Generality is related to notion of purpose which is the extent to which an ontology can or is intended to be reused in a range of different situations. Uschold [42] identifies three main categories of use for ontologies through his work on the Unified Methodology. Firstly, ***communication*** between people for which an unambiguous but informal ontology may be sufficient. ***Inter-operability*** among systems achieved by translating between different modelling methods, paradigms, languages and software tools; here the ontology is used as an interchange format. Lastly, ***systems engineering benefits*** which encompass *re-usability, knowledge acquisition, reliability and specification* and which often rely on a more formal representation of the ontology.

2.4.3.2. Formality and Language Expressivity

Formality captures various dimensions like the level of richness of the internal structure; expressivity and formality of the language used, if any; the type of conceptualization structure and computational capability. Formality as proposed by Uschold [42] is viewed on a continuum from *highly informal;* expressed loosely in natural language, *structured informal;* expressed in a restricted and structured form of natural language to reduce ambiguity; *semi-formal*; expressed in an artificial formally defined language and *rigorously formal*; meticulously defined terms with formal semantics, theorems and proofs of such properties as soundness and completeness. This formality continuum is reflected in a slightly different manner by Gomez-Perez [45] who presents that ontologies belong to one of the following categories, based on the richness of their internal structure: controlled vocabularies, glossaries, thesauruses, informal hierarchies, formal hierarchies, frames, ontologies with value constraints and ontologies with generic logical constraints. Ontologies can be made up of several components: Property, Term, Instance, Concept, Logical Definition, Textual Definition and Instances. Roussey et al. [44] propose that according to which of these components is used, four types of ontologies exist.

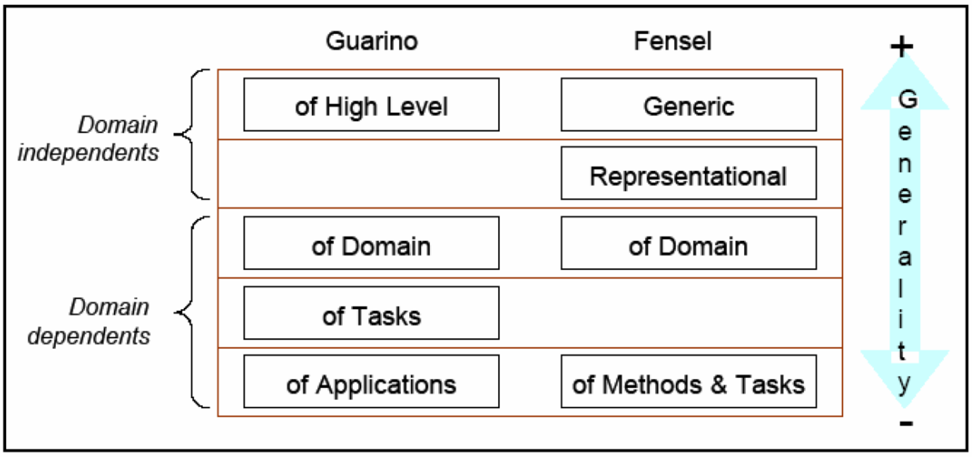
While all the above classification schemes differ in terminology used to describe types of ontologies it can be agreed that all three could be drafted along a formality continuum; the defining dimensions being the level of knowledge representation languages and sophistication of the representation mechanisms used.

2.4.3.3. Generality: Scope of the Ontology and Granularity

Generality, concisely encompasses multiple classification sub-dimensions as cited in literature like the ***scope of the ontology*** and ***granularity***. Scope of the ontology refers to subject of the conceptualisation to be modeled which could be a specific application or viewpoint, a specific domain, or could be domain independent and applicable to many domains. Fonseca et al. [46] outline that ***granularity*** falls along a spectrum of either being coarse-grained or fine-grained. A coarse-grained ontology consists of a minimal number of axioms and is intended to be shared by users that already agree on a conceptualization of the world. A fine-grained ontology has a large number of axioms and requires a very expressive language. The proposal is that the user should move incrementally from coarse to fine-grained ontologies, with the creation of more detailed ontologies being based on more high-level ontologies, so that each new ontology level incorporates the knowledge present in the immediate higher level. The new ontologies are more detailed given that they refine general descriptions of the level from which they inherit.

Roussey et al. [44] explores ontology classification based on scope of the ontology and granularity and provides a concise ontology classification framework which extends beyond the work performed by Guarino [36] and Fensel [47]. This is depicted in Figure 1 which is followed by a discussion of each type of ontology under Roussey et al.'s [44] classification system.

Figure 1: Roussey et al. [44] presents a combined ontology classification system using Generality



Local or Application Ontologies are specializations of domain ontologies where no consensus or knowledge sharing has been achieved [36][44]. This ontology represents a single viewpoint of a user or developer. This ontology is the combination of a domain ontology and task ontology, generally corresponding to roles played by the domain entities when executing an activity. The task ontology contains knowledge to achieve a task, with the domain ontology describing the knowledge where the task is applied.

Domain Ontologies are only applicable to a domain with a specific view point [44][47]. The view point is based on how a group of users conceptualize and visualize some specific phenomenon. They describe the vocabulary related to a generic domain (for example, information systems or medicine) by means of the introduced concepts of higher-level ontologies.

Task Ontologies describe the vocabulary related to a generic task or activity and contain the knowledge to achieve a task (e.g. development or sales), by means of specialization of higher-level ontologies [44][47].

Core Reference Ontologies are a standard that is used by different groups of users that is often built to catch the central concepts and relationships of the domain. This type of ontology is linked to a domain but it integrates different viewpoints related to specific groups of users.

General Ontologies (also referred to as generic or common ontologies) are not dedicated to a specific domain or field [36][44]. They contain general knowledge over a broad area which is reusable in distinct domains, for example, vocabulary related to things, events, time and space.

Upper Ontologies (also referred to as high-level or foundational or top-level ontologies) are generic ontologies that are valid for a variety of domains but are independent of any specific domain or problem [36][47]. They define basic concepts and notions like space, time, states, objects, relations, events and processes. Their purpose is to unify criteria between large communities of users.

Gomez-Perez et al.[45] suggest further specializations like ***Knowledge Representation ontologies, Domain Task ontologies*** and ***Method ontologies*** but the set above is consistently introduced in the literature.

2.4.4. Ontology Representation Languages and Tools

Uncertainty over what can be classified as an ontology remains apparent in research. Informal means by which to specify ontologies can take the form of UML class diagrams, entity relationship models and semantic nets [35][33]. Formal ontology description generally takes the form of an ontology description language. Ontology description languages provide a mechanism to encode the vocabulary of a domain.

Standard ontology languages for the Semantic Web community include the **Resource Description Framework (RDF)**; a standard model for data interchange on the Web, **RDF Schema (RDFS)**; a general purpose language for representing simple RDF vocabularies on the Web and the **Web Ontology Language (OWL)**; a Semantic Web Language which builds on RDFS and provides a language for defining structured, Web-based ontologies. RDF, RDFS and OWL are now widely accepted and established with stable tools for creating and managing ontologies. These Semantic Web languages are inherently built with a graph-based open data model and can thus support integration from different data sources and applications. Further to this, they implement formal knowledge representation; enabling automatic processing and inference about data.

Islam et al. [54] reinforces the language choice of the Semantic Web community through the evaluation of XML-based languages as demonstrated in **Table 8,** noting OWL has W3C support and strong expressive power. A good ontology language should provide a range of constructs to specify the ontology and should be supported by well-known ontology tools. To note, XML-based languages are the natural choice over non-XML-based languages as they address interoperability; a core requirement of the Semantic Web.

Table 8: Evaluation of XML-based ontology languages from Islam et al. [54]

Language	**W3C Support**	**Expressive Power**	**Editors**	**Other Supporting Tools/Languages**
SHOE	No	Low	RFEdit	PIQ SHOE Search
RDF/RDFS	Yes	Medium	Protégé, NeOnToolkit, DOE, pOWL, Swoop, TopBraid	Jena, SPARQL, RD-QL
OIL	Yes	Low	OilEd	Fact
DAML + OIL	Yes	Low	OntoEdit Ontolingua, WebODE, IsaViz	DQL, Pellet, Racer
OWL	Yes	Strong	Protégé, DOE, Altova Semantic Works, WebODE, Swoop, Morla	Jena, OWL-QL, Fact++, Pellet, RacerPro

The discussion below focuses on the standard ontology languages used by the Semantic Web community.

2.4.4.1. RDF and RDFS

The Resource Description Framework (RDF) is a key knowledge representation framework for the Semantic Web. The RDF extends the linking structure of the Web to use Uniform Resource Indicators (URIs) to name the relationship between things as well as two ends of the link; referred to as a "triple". The underlying structure of any expression in RDF is a collection of triples, each consisting of a subject, a predicate and an object. The subject denotes the resource, while the predicate asserts a relation between the subject and the object. Each node (objects and subjects) and the predicate is uniquely identified by a URI. This mechanism to describe resources is a major component of Semantic Web activity; enabling users to deal with information with greater efficiency and certainty.

The Resource Description Framework (RDF) has features which facilitate data merging even where the underlying schemas differ; allowing structured and semi-structured data to be mixed and shared across different applications. RDF graphs are usually serialized and stored in Extensible Markup Language (XML) although other serialization formats may be used.

While RDF provides a representation framework, RDF Schema (RDFS) provides the vocabulary for specifying ontologies or schemas to structure the information captured in RDF. RDFS allows for description of concept hierarchies and the relations between them, with logical inference being defined by a set of rules. RDFS uses a standard vocabulary for denoting classes (rdfs: Class) and properties (rdfs: Property); including the domain (rdfs: domain) and the range (rdfs: range) of these properties. Class hierarchies are represented by relations such as subclass (rdfs: subClassOf).

While RDFS is the simplest and most scalable XML-based ontology language, it is not expressive enough for many ontological encodings. RDFS is limited to specifying concepts and the binary relationships among those concepts.

2.4.4.2. The Web Ontology Language (OWL)

The Web Ontology Language (OWL) [57][85], on the other hand, is quite a sophisticated language, being based on solvable mathematical logic or what is better termed "description logic". OWL is designed to represent complex knowledge about things, groups of things and relationships between things. OWL takes from the ability of the RDF to clearly state facts and the class-and property-structuring capabilities of RDFS and extends them through incorporation of additional vocabulary to express classes, relationships between classes, cardinality, equality, characteristics of properties, annotations and instantiation of classes. RDFS and OWL have different capabilities which facilitate the creation of detailed content descriptions and articulation of logical inference. Horrocks et al. [55] point out that the major extension over RDFS is the ability in OWL to provide restrictions on how properties behave that are local to a class.

For example [55]:

- *Owl:Class* is used to represent class; a specialization of *rdfs:Class;*
- *Rdf:Property* is used to represent property and is divided into *owl:ObjectProperty* and *owl:DatatypeProperty* which differentiate properties that apply to objects and to literal antetypes respectively.

There are two versions of the OWL language; OWL 1 and OWL 2. OWL 2, like OWL 1, was developed by the World Wide Web Consortium (W3C) Web Ontology Working Group and is an extension and revision of OWL 1 [94]. OWL 1 has three major sublanguages which provide for increasing levels of expressivity[1]; with a trade-off to decidability (also known as *efficient reasoning*) [56]. OWL 1 Lite is the most expressive and least decidable sublanguage; OWL 1 Full the least expressive and most decidable. OWL 1 DL remains both expressive and decidable. While OWL 1 has been successful, certain issues within its design called for revision and the development of OWL 2.

OWL 1 presented expressivity issues, problems with the relationship between the two normative syntaxes, an inability to serve meta-modelling (i.e. using classes and properties as individuals) and design choices which led to deficiencies in the definition of OWL species [85]. While OWL 1 DL is the most expressive of the sublanguages, it still lacked particular constructs necessary for modeling complex domains. OWL 2 mitigates both the expressivity and design choice challenges of OWL 1; marking the major difference between the two. OWL 2 is not only more expressive but more decidable than OWL 1 [85][94]. Several extensions made to OWL 2 have led to an improved syntax through the introduction of Functional-Style syntax. Lastly, the meta-model problem was resolved in OWL 2 by specifying the structure of the ontologies using the Meta-Object Facility (MOF); a well-known meta-language [85][94].

OWL 2 has two major dialects, OWL 2 DL and OWL 2 Full [94]. Further to this, OWL 2 has three sublanguages which have favourable computational power and are easier to implement [94]. While they have restricted expressivity, they are sufficient for a variety of applications. OWL 2 EL was designed with large biomedical ontologies in mind and accommodates complex structural descriptions and huge numbers of classes. OWL 2 QL is designed to enable easier access to data stored in databases and can serve as a translational layer for relational database management systems. OWL 2 RL is aimed at applications that require scalable reasoning without sacrificing too much expressive power.

OWL Syntaxes

OWL 2 ontologies are typically stored using the RDF/XML syntax [57] so that they may be exchanged among tools and applications; however, there are several different syntaxes for creating ontologies. RDF/XML syntax is verbose but is the most widely supported OWL syntax. Another common syntax is the Manchester syntax [57] which is specifically designed to be easily readable by non-logicians. As mentioned earlier, several extensions made to OWL 2 have led to an improved syntax through the introduction of Functional-Style syntax. The Functional-Style syntax is concise and allows for ease of specification. Less commonly used are Turtle [57] and OWL/XML syntaxes [57].

Open World Assumption

The Open World Assumption (OWA) is a key tenet of the Semantic Web Languages, distinguishing it from the likes of SQL and Prolog. OWA caters for the fact that people will extend models in the future; allowing for incomplete information by default. OWA assumes that when a statement cannot be proved true; the default specification is not false but unknown. The Closed World Assumption assumes

[1] **Expressivity of a language** refers to the breadth of ideas that can be represented and communicated in that language. The design of languages and formalisms involves a trade-off between expressive power and analysability. The more a formalism can express, the harder it becomes to understand what instances of the formalism say. Decision problems become harder to answer or completely undecidable.

information is complete and would require an input of false in the case where a statement is not proved true. Incomplete and fragmented information is common with respect to the Semantic Web and ontologies, hence the OWA is very important in upholding these constructs.

Reasoning with OWL

Description logic (DL) provides the underlying formal framework for both OWL and RDF; allowing for inference to be performed on the knowledge described with OWL. DL describes domains in terms of concepts (classes), roles (properties and relationships) and individuals; providing decision procedures for key problems (e.g. satisfiability, subsumption etc.). The DL knowledge base is constructed through the combination of a Terminological Box (TBox); a set of "schema" axioms which are declarations about concepts, and the Assertional Box (ABox); a set of "data" axioms which are facts about objects and statements about these facts.

An example illustrates the concept of both:

A TBox
{Doctor → Person, HappyParent ↔ Person ^ [hasChild](Doctor v {hasChild}Doctor)}

An ABox
{John → HappyParent, John → {hasChild} Mary}

An OWL ontology is equivalent to a DL Knowledge Base. In order to produce additional inferences in the ontology a reasoner is used to apply rules of OWL to the statements contained in both the TBox and the ABox.

Table 9: Description Logic Example

OWL Syntax	DL Syntax	Example
Type	$a : C$	John : Happy-Father
Property	$\{a,b\} : R$	{John, Mary}: has-Child

Table 9 specifies well defined semantics through specifying individuals (John, Mary), concepts (Happy-Father) and roles (hasChild).

2.4.5. Ontology Editors

Ontology Editors provide a user-friendly interface for ontology development. A good ontology editor should provide an easy user interface for ontology creation, be supported by well-known ontology tools and provide support for major ontology languages. Islam et al. [54] point out that most of the current ontology editing tools have evolved from research projects as a proof-of-concept for some research idea. Protégé, OntoEdit and IsaViz provide powerful desktop interfaces for ontology editing. Further to this, these tools provide support for well-known ontology languages like RDF(S) and OWL. Other editors such as WebODE, SWOOP and NeOnToolkit provide support for multiple users to collaborate and create an ontology. Some editors like WebODE, DOE and OntoEdit recommend that a particular methodology for ontology engineering be employed. While the NeOn toolkit has a sleek and intuitive UI, it lacks support for some of the most recent features of OWL, including the ability to specify keys

[57]. The choice of the particular ontology editing tool depends on the particular requirement of the Semantic Web application for which the ontology is being developed.

Ontology editing tools have been updated to accommodate the improvements of OWL 2. For instance, the Protégé ontology editor has been extended in its newer versions to support the additional constructs of OWL 2. Domain ontologies are often developed using Protégé which serves as a rapid development environment [43]. The Protégé OWL plugin provides an open testing framework to allow code to be executed at any time. Protégé has many useful features, completely supports all OWL features and allows for saving an ontology in all of the various OWL syntaxes, including several convenient visualization tools. Notably, the Protégé ontology editor is free and is known as "the leading ontological engineering tool" [58]. Protégé also implements reasoners to facilitate additional inference; these include Pellet, RacerPro, Fact++, HermiT and KAON2.

2.5. Ontology Engineering

A formal methodological approach is advisable for ontology development of complex or large-scale domains. Ontology engineering methodologies have been described below to discern which may be applicable to developing an ontology which captures the complexity of share evaluation in the investment domain.

2.5.1. Design principles for Ontology Engineering

Gruber [38] proposed a broad set of design criteria for ontologies to encourage knowledge sharing, the development of shared conceptualization and enhanced interoperability. The key criteria as proposed by Gruber have been detailed below.

- **Clarity** requires that the intended meaning of the defined terms is communicated effectively with definitions clear and unambiguous, whether expressed in natural language or formally encoded. Examples and underlying assumptions should be stated where possible.
- **Consistency and Coherence** is achieved by ensuring that definitions of the defined terms and their inferred meanings via axioms are logically consistent. The ontology should be both internally consistent; with circularity avoided, and externally consistent; with terms conforming to common usage.
- An **extendable and reusable** ontology is one in which new terms for special uses can easily be defined without requiring the revision of existing definitions. This requires a balance in specificity: specific enough to perform the required tasks; not so specific to detract others from using it.
- **Minimal encoding bias** ensures that where possible, the model should be specified at the knowledge level as opposed to a particular symbol level encoding which requires further interpretation.
- **Minimal ontological commitment** requires that the ontology should make as few claims as possible about the domain of knowledge, allowing parties committed to the model the freedom to specialize or instantiate the ontology as needed. This means that one should be very careful to introduce several terms that mean the same thing; instead define the key underlying term and reuse it to define other terms.

2.5.2. Methodological Decisions

While many of the broader design principles for ontologies are agreed among the computer science community, this consensus does not extend to ontology engineering methodologies. Each methodology must be evaluated to identify the most suitable for the type of ontology to be developed. A few methodological guidelines from research have been outlined below and are applicable to most ontology engineering methodologies.

- **Reuse existing** ontologies as much as possible in your ontology [48].
- **The middle-out approach** rather than top-down or bottom-up should be employed in choosing which terms to define first. The middle-out approach avoids the high level of detail and issues arising from the bottom-up approach and has been proven successful over many years [49].
- **Concentrate on underlying ideas first** and ignore terms to reach agreement when terms are used ambiguously [49]. First define each idea, developing meaningless labels for each, then decide on the most important ideas and finally the terms.
- **It is imperative that consensus finds its way into the development process among relevant domain experts** [33]. The requirement for consensus places ontological development success at risk given that human interaction is required; investment in earlier phases of ontology-design is intensive [33].
- **A means by which to communicate with non-technical users must be found** [33] [35]. Many do not understand the technical notations used for modelling ontologies nor do they understand the intention of the designer of the ontology [35]. Ontological engineering must take into account an "independent business record" to design against; a "business conceptual model".

2.6. Ontology Engineering Methodologies

This section compares the main ontology engineering methodologies, namely: the Unified Methodology, MethOntology and the Unified Process for Ontology Building.

2.6.1. Methodology 1 – Unified Methodology

The Unified Methodology proposes different guidelines for developing ontologies depending on how they are classified based on three key dimensions: purpose, level of formality and scope. Through a critical comparison of the TOVE and Uschold and King's development methodologies, Uschold [42] reveals that a better framework for methodological choices needs to be in place. With this in mind, the Unified Methodology lays a framework which identifies and separates steps and techniques which have general applicability to all ontologies and those that do not. The framework starts with **identifying the purpose of the ontology** through identifying the target users, identifying general motivating scenarios and competency questions and producing a user requirement document. Following this, the **level of formality** should be decided. In some cases, both a formal and informal ontology may be required to satisfy both technical and non-technical users. Once the purpose and formality level have been defined, **the scope must be defined** through a set of concepts and terms covering the domain that the ontology must represent. The scope can be clarified through the creation of detailed story problems that arise in the applications and possible solutions to these problems. In addition, the competency questions should be completed to specify the full reasoning requirements of the ontology. The final output of scoping should be a set of concepts and terms that must be included in the ontology, whether or not structured in some way. The set of terms and concepts outlined are then formalized in the build phase.

Four possible approaches for the build phase may be followed depending on the formality required, intended purpose and what has been done in prior stages. **Approach 1** describes a hacker's approach; adequate for prototyping small or simple ontologies, and entails one skipping the scoping exercise as described above to directly engaging with an ontology editor to define terms and axioms. **Approach 2** suggests that one has gone through the scoping exercise; and from this point begins a formal encoding; it may provide a better result for small or simple ontologies. **Approach 3** suggests a more diligent option through creation of a complete intermediate document containing terms and definitions in a structured form of natural language. The creation of this informal ontology could serve several purposes: (1) be the end result, where no formal encoding is required (2) serve as a specification for the formal code and/or (3) be documentation for it. **Approach 4** suggests that no intermediate document be developed and extends Approach 2. One proceeds by identifying formal terms from a set of informal terms, following which the axioms and definitions are specified.

Further work remains to refine the Unified Methodology and where possible, merge this methodology into a coherent framework.

2.6.2. Methodology 2 – MethOntology

A complete ontology development process was initially proposed in [48] and updated in [50] by Fernandez et al. MethOntology presents a structured method to build ontologies from scratch. MethOntology seeks to describe the major tasks to be performed as described below and provides a suggested life cycle of the development process.

Specification seeks to produce an ontology specification document which records the following: (i) the purpose of the ontology (ii) level of formality of the implemented ontology i.e. will terms and their meanings remain in natural language or need to be encoded in a rigorous formal language (iii) the scope, including the terms to be included and its granularity.

Knowledge acquisition is an independent activity in the ontology development process and occurs concurrently with other phases. Most of the knowledge acquisition should occur simultaneously with the requirements specification phase and decreases as the ontology development process moves forward. Experts, books, handbooks, figures, tables and other ontologies are recommended to be used in conjunction with active techniques such as brainstorming, interviews, and formal and informal analysis of texts. Knowledge acquisition should enlist various techniques to create and elaborate on the first glossary with terms that may be potentially relevant and then to refine this, adding and removing terms and grouping concepts.

Conceptualization involves structuring the domain knowledge through a conceptual model which is used to describe the problem and how it can be solved. The conceptual model is expressed as a well-defined deliverable: a complete Glossary of Terms which contains all usable domain knowledge with terms and their meanings identified and gathered. Terms in this instance encompass concepts, instances, verbs and properties.

Integration assists with speeding up the construction of the ontology through re-use of definitions already built into other ontologies instead of starting from scratch. This involves evaluating meta-ontologies for fit with the current conceptualisation and seeking out libraries of ontologies with terms coherent with the terms identified per the conceptualisation. An integration document should be created which contains the name of the term in the conceptual model, the name of ontology from which it will take its definition and the name of the definition and its arguments within that ontology.

Implementation entails the codification of the ontology into a formal language.

Evaluation encompasses technical judgement of the ontology. This phase encompasses both verification; the technical process that guarantees the correctness of the ontology, and validation; the process that guarantees the ontology represents the system it is supposed to. MethOntology proposes documenting the evaluation techniques employed, the errors found and the sources of knowledge used in evaluation.

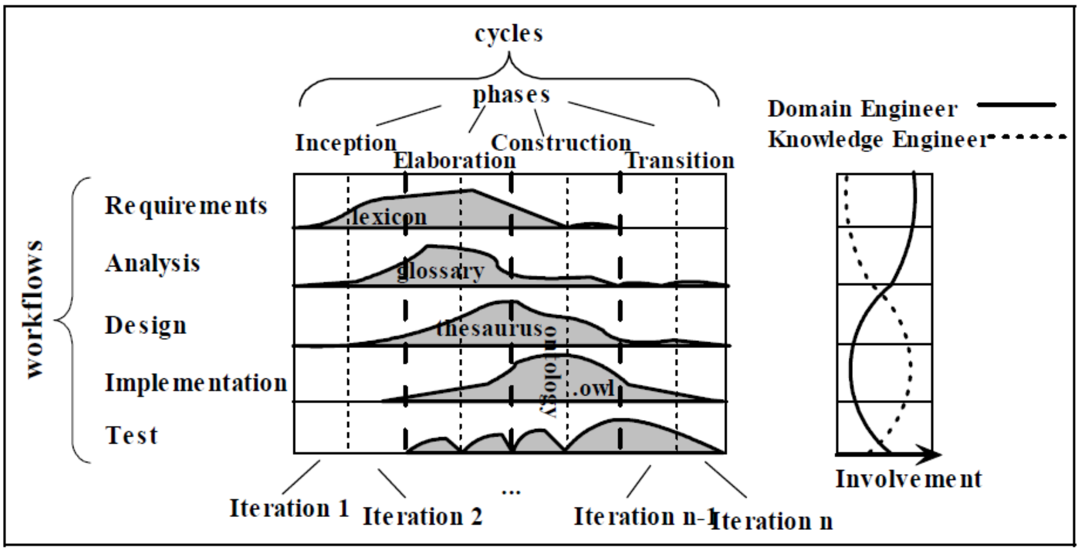
Consensus on **documentation** of ontologies has not been reached. One can go further than the traditional documentation for an ontology through documentation of the development process. MethOntology attempts to break the cycle of poor documentation during ontology development, through incorporation of the documentation activity into every phase of MethOntology.

2.6.3. Methodology 3 – Unified Process for Ontology Building

While the engineering methodologies proposed above are focused on ontology design, Ogundele [85] illustrates the use of UPON (Unified Process for Ontology Building) in the design of semantic Bayesian networks. UPON is often proposed as a methodology for the design of ontologies given that it is derived from one of the most widespread and accepted software engineering process, the Software Development Unified Process (UP) [52]. The UP and the UPON are distinguished from other software and ontology engineering processes respectively given their *use-case driven, iterative and incremental* nature [52]. UPON aims to produce an ontology with the purpose of serving its users, both human and automated systems and use-cases drive exploration of all aspects of the ontology. The nature of the process is both *iterative;* each activity is repeated to concentrate on different parts of the ontology being developed and *incremental* given that each cycle allows for a more detailed, refined and extended ontology to be developed. The incremental nature of the UPON methodology leads to identification of relevant terms in a lexicon; which are further enriched through the addition of definitions, yielding a glossary; the addition of basic ontological relationships yields a thesaurus and a final formalization produces the reference ontology.

UPON consists of cycles, phases, iterations and workflows. **Each cycle** consists of **four phases** (*inception, elaboration, construction* and *transition*) with the result of each being a new version of the ontology. **Each phase** is further **sub-divided into iterations**. During **each iteration**, **five workflows** take place (*requirements, analysis, design, implementation* and *test*). More than one iteration may be required to complete each of the four phases. The domain expert provides their contribution in the early workflows; *Requirements* and *Analysis*; and partially during *Test.* The knowledge engineer is mainly focused on the *Design* and *Implementation.*

Figure 2: UPON methodology broken down into cycles, phases, iterations and workflows



The ***Inception Phase*** is mostly concerned with capturing *requirements* and partly performing some conceptual *analysis*. Implementation and test are not performed.
The ***Elaboration Phase*** entails *analysis* and the loose identification of the fundamental concepts. Some *design* effort may occur, with *implementation* a small skeletal blueprint of the ontology.
The ***Construction Phase*** is pervaded by design and implementation workflows, with some further analysis to identify further concepts to be added to the ontology.
The ***Transition Phase*** entails heavy testing with the ontology eventually being released.

The **workflows** are carried out through the **phases** of the ontology with shift in attention from requirements to design to implementation. Each phase renders a partially complete ontology, with the early phases focusing on establishing the requirement and the later phases resulting in incremental ontology release. The workflows have been detailed below.

The requirements workflow entails specifying the semantic need and knowledge to be encoded in the ontology. The modeler and domain experts must agree on the fragment of reality to be modelled. There are six steps suggested to achieve this goal:

(i) Determining the domain of interest and scope
(ii) Defining the purpose
(iii) Writing a storyboard
(iv) Creating an application lexicon
(v) Identifying the competency questions
(vi) Identifying the related use cases

The analysis workflow entails refining and organizing the ontology requirements identified in the requirements workflow. Re-use of existing resources is encouraged and includes identifying existing ontologies, lexicons and documents. This step entails defining a reference lexicon to encompass the key concepts and their definitions to be used in the ontology. Domain experts are required to agree on the definition of concepts. These definitions should contain precise references or mentioning the author of that definition. A first version of a glossary of concepts of the domain of interest has to be built merging the application lexicon (from the domain experts) and the domain lexicon (from the existing resources). Merging of the two lexicons will pinpoint two major areas: the intersection area and the disjoint area.

The glossary should include all concepts that are in the intersection area with consideration for inclusion of the remaining concepts in the disjoint area by domain experts.

The design workflow entails the transition from lexicon terms to concepts and refines the concepts in the analysis workflow through the addition of relationships. The concepts and relationships between concepts are developed into a hierarchical model. To develop the hierarchy a *middle-out* approach is adopted which entails identifying the salient concepts and then generalizing and specializing them. This approach is considered to be the most effective as concepts 'in the middle' tend to be more informative about the domain. The final output of the workflow is a semantic model that is ready for implementation with an ontology language.

The implementation workflow entails the formalization of the ontology in a language and its implementation in terms of components. Components implement concepts from the design workflow and follow the established grouping into packages. While components can be written in many languages and notations, a long standardization effort has resulted in the Ontology Web Language (OWL) being the main candidate for encoding an ontology to be used on the Semantic Web.

The test workflow verifies that the ontology correctly implements its requirements and validates that it reflects the intended reality it was designed for; ensuring that the coverage is sufficient over the application domain and that the ontology can be used to answer competency questions.

2.6.4. Critical Comparison of the Ontology Engineering Methodologies

The three ontology engineering methodologies described above are evaluated below. A prior review of these methodologies was performed by Iqbal et al. [53]. Iqbal et al. [53] developed a criterion set for comparing ontology engineering methodologies to evaluate their maturity and wide acceptability; this has been adopted for this research paper's evaluation needs. An extract of their comparison is presented in **Table 10** below. The criteria were established through review of related literature and observing trends and needs which have evolved over the years in the field of ontology engineering. The defined criteria allow for a quick evaluation and understanding of different methodologies but also aid in the selection of the correct methodology given project needs, preferences and priorities. The first four aspects of the criteria focus on the high-level details of the ontology whilst the remaining four aspects evaluate the technical side of the methodology. Iqbal et al. [53] demonstrate that there is no completely mature methodology.

Table 10: Comparison of Ontology Engineering Methodologies from Iqbal et al. [53]

Set Criteria	Unified Methodology	MethOntology	UPON
High Level Details			
Type of development	Stage Based	Evolving Prototype	Evolving Prototype
Support for collaborative construction	No	No	No
Support for reusability	Yes	Yes	Yes
Support for interoperability	No	No	No
Technical Details			
Degree of application dependency	Application Independent	Application Independent	Application Independent
Life cycle recommendation	No	Yes	Yes

Strategies for identifying concepts	Middle-out Strategy	Middle-out Strategy	Middle-out Strategy
Methodology details	Some details	Sufficient details	Some details

Based on the framework provided by Iqbal et al. [53], all three methodologies are application independent and support reusability. All explicitly adopt a middle-out strategy for identification of concepts. A middle out strategy entails identifying the most important concepts first, before generalizing or specializing for other concepts. MethOntology and UPON both propose a life cycle implementation and follow an evolving prototype model for ontology building which is the best choice where requirements are unclear and need to be refined over time. The Unified Methodology proposes a stage-based approach which is suitable in cases where the purpose and requirements are clear. MethOntology remains the only ontology with sufficient details of techniques employed in building an ontology which lends to easy adoption of this approach and a more widespread user base. This criterion should not eliminate other methods like UPON which provides significant detail on many aspects within the ontology development process.

Further to the above evaluation approach, De Nicola et al. [52] proposes another way to evaluate ontology engineering methodologies based on how sufficiently the methodology addresses each development phase and process. This has been summarized below in Table 11.

Table 11: Evaluation of Ontology Engineering Methodology from De Nicola et al. [52]

Development Phase	**Processes**	**Unified Methodology**	**MethOntology**	**UPON**
Project Management Processes	**Project Initiation**	Partial	No	No
	Monitoring and Control	Partial	Partial	No
	Quality Management	No	Partial	No
Pre-development	**Environment Study**	No	No	Partial
	Feasibility Study	No	No	No
Development	**Requirements**	Partial	Yes	Yes
	Design	No	Yes	Yes
	Implementation	Yes	Yes	Yes
Post-development	**Installation**	No	No	No
	Operation	No	No	No
	Support	No	No	No
	Maintenance	No	Partial	Partial
	Retirement	No	No	No
Integral Processes	**Knowledge Acquisition**	Yes	Yes	Yes
	Evaluation	Yes	Yes	Yes
	Configuration Management	No	Yes	No
	Documentation	Yes	Partial	Yes
	Training	No	No	Partial

The Unified Methodology fails to address many of the integral processes necessary to develop an ontology while the UPON and MethOntology development methodologies focus on the core activities of ontology building with the integral processes well addressed. Peripheral activities like project management and many other pre- and post-development activities are largely out of scope for these methodologies.

2.7. Ontologies in Finance

This section provides context on the application of ontologies in Finance. Ontologies are used to address two key research problems that are common across the financial domain and which are relevant to the more specific problem of share evaluation. The two research problems are: Web Content Management, and Automated Analysis and Decision Support which have been discussed below. This research traverses both research problems within the context of share evaluation.

2.7.1. Web Content Management

Web content management is focused on ensuring that data is ready for meaningful analysis through collection, organization, and integration of heterogenous data sources which have been semantically labelled. This issue has become increasingly important to financial professionals given the unending proliferation of complex, voluminous information and data from a variety of heterogenous data sources on the Web [59]. Ontologies are able to stipulate the semantics through articulation of concepts and the relationships between these concepts; making data understandable to machines such that the data may be extracted and processed for automated analysis and decision support.

Web content management has received considerable attention in industry and academic research, and through this Semantic Web technologies have achieved a certain degree of maturity; providing a consistent and reliable basis for organisation, manipulation and visualisation of financial data. In the last few years several finance-related standards, ontologies and taxonomies have been developed; these are discussed below.

2.7.1.1. Web Content Management in Industry: XML Standards for Financial Contents and Business

XML standards have arisen through consortiums and professional bodies to articulate concepts and the relationships between them for common business processes and transactions. A key standard for financial decision makers is the eXtensible Business Reporting Language (XBRL), a core taxonomy which formalises concepts pertaining to financial statements and the relationships between them [60] [61]. In 2008 the US Securities and Exchange Commission (SEC) mandated all tier-1 publicly traded companies to submit their official financial statements in XBRL [62]. In July 2018, the Companies and Intellectual Property Commission (CIPC) in South Africa followed suit; public listed entities who are required to file their annual financial statements with their annual return must do so via Inline XBRL (iXBRL). iXBRL allows for documents to be viewed on standard web browsers as opposed to an unstructured PDF format. The implementation of iXBRL has meant that financial statement data is now formally encoded and accessible in a machine-readable format to users of financial statements. Digital collection and organisation of financial statement information from a publicly available source is now possible.

While XBRL focuses on content structure and provides a rich vocabulary of terms for content classification, there is need for its extension through deeper semantics and representation to describe complex knowledge [43]. The fact is, semantic differences across XBRL filings remain and pose a challenge for investors seeking to make inter-firm comparisons using XBRL data [63]. Li et al. [64] provide a clear example where XBRL is not sufficient: trading in financial instruments may be an investing activity for certain companies like Coca Cola while for others like Goldman Sachs it may be an operating activity; this is not yet clearly stipulated via XBRL. These distinctions need to be made explicit through deeper, rich semantics. XBRL has scratched the surface of what is needed for precise search and automated data consumption [64][96]. Semantic Web technologies are continuing to alleviate these problems through the development of models that procure combinations of smaller pieces of domain specific content which fit the desired need of the particular user and allow for reusability. Pease et al. [65] anticipate that communities will adopt domain specific ontology standards with deeper semantics to facilitate the exchange of information such that there will only be a handful of upper level ontologies to which all applications will be compliant; allowing for mutual interoperability.

2.7.1.2. Web Content Management in Academic Research: Upper Level Ontologies

Upper level ontologies are ontologies which consist of general terms that are common across all domains and serve to support broad semantic interoperability among a large number of domain-specific ontologies. Business processes like transaction processing, risk management, compliance, general ledger accounts management and reporting rely on information about the same things but have their own specific ways of looking at information. An upper level ontology can resolve integration challenges inherent in business processes which involve a large number of businesses exchanging data in a range of formats and message protocols [35]. A few of the latest upper level ontologies developed through research consortiums and groups are discussed below.

The **Suggested Upper Merged Ontology** (SUMO) is an upper level ontology which was created through combining several ontologies. Examples of the extension of this upper level ontology for more domain specific ontologies are **SumoF**, which focuses on banking and investment finance, and **SumoS** which is capable of representing e-commerce services. **BORO** (Business Object Reference Ontology) was developed as a basis for facilitating, among other things, the semantic interoperability of enterprises' operational systems. On the other hand, **TOVE** (Toronto Virtual Enterprise), developed by the Enterprise Integration Laboratory from Toronto University, is an ontology built for enterprise integration that each agent in a distributed system can jointly understand and use. The **XBRL Ontology Specification Group** developed a set of ontologies for describing financial and economic data in RDF for sharing and interchanging data. This ontology is becoming an open standard means of electronically communicating information among businesses, banks, and regulators. The **consortium DIP** (Data Information and Process Integration) developed an ontology for the financial domain which was mainly focused on describing Semantic Web services in the stock market domain. The **Finance Ontology** is an upper level ontology used to represent knowledge in the financial services domain including the financial statements. Economic **R**esources, economic **E**vents, and economic **A**gents (**REA)** is used for reasoning about accounting concepts and business enterprise phenomena [66].

To some extent many of these ontologies have been eclipsed by the **Financial Industry Business Ontology (FIBO)** which is larger, well-documented and has many similar terms [67]. This has been discussed below.

The Financial Industry Business Ontology (FIBO)

The Financial Industry Business Ontology (FIBO) is one of the more widely used upper-level ontologies developed for the Financial Domain. FIBO is understandable to both business and technical people alike, is extremely well organized and has exceptional visualizations [68]. FIBO is two things: a business conceptual ontology and an operational ontology delivered together. The business conceptual ontology provides a precise visual representation of financial concepts which have been vetted by business subject matter experts. Further to this, the operational ontology is aligned with the business conceptual ontology, however this is a subset and is geared towards practical application [69]. FIBO is an open semantic standard and industry initiative to define financial industry terms, definitions and synonyms using RDF/OWL, RDFS and modeling standards like UML [35]. FIBO arose out of the joint effort of Object Management Group (OMG) and the Enterprise Data Management (EDM) Council which aims to contribute to semantic disambiguation in the financial industry, support further derivation of ontologies and semantic models through use as a reference model and encourage and support the use of formal business definitions in legal documents [70]. FIBO is a modularized formal Business Conceptual model[2] of concepts represented by financial industry terms. The modular nature of FIBO ensures that it is easily re-usable.

The Foundations module of FIBO specifies a set of business concepts which are intended to define general financial concepts which are applicable to most businesses, not just the financial industry. The Accounting Module contained within FIBO Foundations contains ontologies of general accounting concepts including assets, financial assets, and equity and specific properties which allow one to reason with financial data.

The choice of Semantic Web standards has ensured that FIBO is defined formally in a well-known and widely referenced standard and to enable onward mapping and extension. The extension of FIBO is illustrated through the development of the Global Fund Reporting Ontology (GFRO). The GFRO aims to advance research by illustrating the application and extension of FIBO for improved reporting capabilities over a broad subset of financial instruments; specifically, bonds and equities. In this case, FIBO is adapted to bridge the gap between domain specific databases and the FIBO ontologies. Another example is the Financial Regulation Ontology (FRO) which aligns and extends FIBO and Legal Knowledge Interchange Format (LKIF) and defines semantic rules that implement laws and regulations of Dodd-Frank and the Investment Adviser Act.

2.7.1.3. Web Content Management in Academic Research: Lower Level Ontologies

Web content management is not reserved for upper level ontologies; lower level ontologies are developed to address semantic inadequacies of the Web through the articulation of concepts and concept relationships such that specific applications may more easily parse, organise and use the relevant collected data. There are a multitude of lower level ontologies that deal with web content management. For example, Du et al. [32] explore an ontology-based framework for effectively managing big financial data while an Algeria case study [71] depicts an ontology for financial investments which is intended to be used for integration with a financial intelligence system. On the other hand, Shan [72] developed an ontology which captures domain knowledge about financial news; specifically, the primary classes of news, classes of financial markets participants, classes of financial instruments, and relations between these classes.

Web content management aims to improve the quality of data available which is as an essential factor for decision performance [60]. Given the high frequency, high diversity and dependency of financial data, traditional or manual methods to address data quality problems are ineffective. Data can suffer

[2] A **business conceptual model** uses the full expressive power of the chosen notation to formally define items in the domain of discourse, without taking application technical constraints into account. In contrast an operational ontology is constrained to operate within the parameters of a specific semantic application which will contain a sub-set of the constructs in the business conceptual ontology.

from quality problems like terminological ambiguity, conceptual inaccuracy, missing data, unreliable data, inconsistent representation and incomplete domain. Du and Zhou [60] developed their financial ontology of the income statement to illustrate synergic semantic alignment which entails the “conversion of multiple input data which do not refer to the same object or phenomena to a common object or phenomena [73]”. Kalcheva [61] explores this in further detail with a linked data proposition which seeks to improve the quality of data required by all financial decision-making applications. As pointed out, financial data published online suffers from redundancy; it is duplicated across many online sources [61]. That being said, the data is presented differently across sources in spite of high regulation in the financial domain.

Existing application-specific ontologies have been developed for a variety of application-specific problems or specific issues related to web content management. These lower-level ontologies cover diverse and specific domains; their re-use is often hindered due to poor documentation. It is evident that recommended design practice to re-use and extend existing upper level ontologies is still nascent in lower level ontology research.

2.7.2. Automated Analysis and Decision Support

Ontologies are not limited to addressing semantic inadequacies through web content management but provide a mechanism to represent how experts reason with data. Ontologies may be used to formalize and represent expert knowledge such that machines can provide automated analysis and decision support. Automated analysis and decision support have gained focus in research for many application areas within Finance given that the volume of data and range of data points included in the decision process is expanding exponentially. The increase in the volume of data due to economic globalization and the rapid evolution of information technology means that manual analysis is becoming increasingly costly, time consuming and often, beyond our human limits [74][61]. In addition, financial professionals are seeking to further enrich their current models with more types of data and more granular data. Examples include Tweets from Twitter and news headlines which are almost impossible to analyse without computers. Financial professionals are in critical need of automated approaches to ensure effective and efficient utilization of large amounts of financial data to facilitate investment decision making [74]. In addition, financial professionals are becoming increasingly aware of bias and human error inherent in manual decision-making processes and are leaning towards automated processes to improve the robustness of financial decision making.

2.7.2.1. Automated Analysis and Decision Support utilizing Ontologies

Financial decision making can be found across a multitude of categories: stock return prediction, portfolio management and optimization, bankruptcy prediction, foreign exchange rate prediction, detection of fraud, trading models and analysis, loan risk analysis and payment prediction, mortgage scoring, real estate assessments and business performance analysis [74]. In most categories there are several examples of ontology-based approaches that have been used to automate the analysis and decision process for such. Kanellopoulos et al. [43] present an ontology for prediction of firms with fraudulent financial statements through the logical evaluation of twelve financial ratios and the use of a decision tree model. Hu et al. [75] modelled rare risk events to evaluate their effect on banking systems through the Banking Event-driven Scenario-oriented Stress Testing (or simply, BESST), which is a non-probability-based approach for modelling and analyzing exceptional but plausible stress testing scenarios without historical data. The work of Chowdhuri et al. [63] addresses the two research problems of web content management and automated analysis in one paper through the creation of the Ontology-based Framework for XBRL-mapping and Decision-making (OFXD) which attempts to resolve interoperability between different XBRL filings for seven financial items and presents how this can then be used for meaningful automated analysis. The seven financial items are used to calculate ratios based on three financial conditions: profitability, financial leverage and liquidity, and operating efficiency. Automated analysis and decision support are often presented in isolation; this work shows the transition from ontology to a decision support system. A further work, the Fundamental Analysis System for Trading ("FAST") which focuses on share evaluation to predict future share performance has been detailed below.

2.7.3. Fundamental Analysis System for Trading (FAST)

Colomo-Palacios et al. [76] present the Fundamental Analysis System for Trading (FAST) that has been developed using semantic technologies and is designed to aid investors in the investment process through a recommendation of the most appropriate long-term investment decision (buy, sell or hold the shares). FAST allows the generation of investment recommendations for a set of companies, using some financial information stored in an ontology format. **FAST** has been detailed in this research to guide the design of a better and more robust ontology and decision model. **FAST** is more understandable than most MCDSS discussed above given that a financial ontology has been clearly defined to support the decision model. This approach has been adopted in this research paper.

FAST utilizes several ontologies and a reasoning tool to reach an investment recommendation. Financial data is organized via a **financial ontology** and stored in a **database repository**, then extracted by a **financial data reader** and processed by way of a **financial calculator** which contains several rules to create the desired financial ratios which are finally written to the **financial reasoning ontology** which is then able to make an inference regarding a long term investment recommendation. The **financial reasoning ontology** is divided between a set of concepts and a set of rules which have been codified and explained. Two recommendations follow from the financial reasoning ontology: (1) whether the company is a good company to invest in and (2) whether one should buy, sell or maintain an investment in the company. The rules linked to these recommendations are as follows:

Medium Term Prediction Rules facilitate the decision as to whether the company is a good company to invest in on a medium-term basis. The rules are as follows:

(1) **PTB Rule:** A rule is triggered if the price-to-book ratio (PTB) of Company X is greater than or equal to the average of the sector
(2) **PER Rule:** A rule is triggered if the price-to-earnings ratio (PER) of Company X is less than or equal to the average of the sector
(3) **PER∩PTB Rule:** If both the PTB and PER rules are triggered, the PTB intersection PER rule is triggered.
(4) **PCFR Rule:** A rule is triggered if the price-to-cash-flow ratio (PCFR) of Company X is less than or equal to the average of the sector and if the PTB of Company X is greater than or equal to the average of the sector.
(5) **PER∩PTB∩PCFR Rule:** If both the PTB intersection PER rule and PRCR rule are triggered, the PTB intersection rule is triggered. If this rule is triggered, that means that Company X is a good company to invest in on a medium-term basis.

The Long-Term Prediction Rule fulfills the objective of comparing the calculated value called "actual share calculated value" (ASCV) against the current price of a share. Depending on this comparison (one is greater than the other), one of the three investment options will be returned (sell, buy or maintain). This rule determines when ASCV differs sufficiently, that is by a margin of more than 10%, from the current share price. A margin is applied given that fundamental analysis doesn't try to calculate a concrete value to compare with the current share price. The margin allows for a band of values within which the ASCV should move. This research takes note of this approach and incorporates a margin of safety through the use of thresholds into the decision rules.

FAST only includes the necessary financial ratios like name, PER, PCFR, PTB, Share Value and ASCV in the inference process while excluding intermediate financial values which are not necessary, the model design in this research follows this approach to alleviate computational burden.

The financial rules used in the system have been developed using Semantic Web Rule Language (SWRL). SWRL combines sublanguages of OWL (OWL DL and Lite) with Rule Markup Language. **FAST** employs a rule hierarchy to reach an investment decision. Each rule serves to eliminate shares which do not meet the criteria for that rule prior to moving onto the next rule, alleviating the burden of weighing up several criteria at once. **FAST** fails to reflect the fact that the decision process is not only a multi-criteria problem but also an uncertain one. SWRL requires very clear rules for inference and does not allow one to reason with multiple variables at once. For this reason, Bayesian networks were selected over SWRL to correctly reflect and reason with the uncertainty inherent in the decision process. Bayesian networks are introduced and discussed in detail below.

2.7.4. Summary

The two research problems presented above, namely web content management and automated analysis are no less important than the other. Existing research papers generally serve to address one of these application problems but there has become a need for research that deals with both: for more decision support systems to be based on well-defined ontologies which allow one to extract knowledge from data with limited human intervention [61]. **FAST** should be seen as state of the art with respect to ontologies employed for automated share evaluation given that it is most relevant to solving the research aims and objectives and closest to the solution that is envisaged in this research.

2.8. Bayesian Networks

Several limitations exist with respect to how information can be processed through ontologies. While ontologies are exceptional at representing organizational structures of large complex domains, their application remains bounded by their inability to deal with uncertainty [34]. It is not always easy to take informal knowledge and state it in formal terms, especially when the knowledge is uncertain. Bayesian Networks are often integrated into ontology-driven models to alleviate this issue. Bayesian networks (*also* known as Belief Networks) provide a representation tool to capture belief relations, that is informal or uncertain knowledge, between a set of variables which are relevant to some problem. Given that uncertainty is an inevitable feature of the investment landscape and information is often incomplete or imprecise, an ontology-driven Bayesian model for uncertain knowledge representation is essential [34].

2.8.1. What is a Bayesian Network?

A Bayesian network is a compact, flexible and interpretable representation of a joint probability distribution. It is a useful tool in knowledge discovery as directed acyclic graphs allow representation of causal relations between variables.

A Bayesian network is constructed in three steps: (1) creation of a qualitative graphical structure to specify the relationships among variables, (2) the incorporation of probability distributions into this structure (to indicate numerical relationships between dependent variables) and (3) making decisions or inferences based on newly available evidence [77]. Research has indicated that the graphical structure is the most important [77].

Bayesian networks are adaptable; they are able to be started off small with limited knowledge about a domain. One does not require a complete knowledge about the instance of the world one is trying to model. A Bayesian network will do as good a job as possible with the knowledge available. As one acquires knowledge, the probabilities in the network will automatically adjust; reflecting the power of Bayesian inference. Bayesian inference is based on scientific and provable estimates of the likelihood of something; an estimate which becomes more accurate as we gain knowledge.

2.8.2. Bayesian Theory

The real world is filled with vagueness and uncertainty; Bayesian probability is well suited for dealing with real world uncertainty where no complete theory is available, it is too tedious or complex to incorporate all the required observations and all the necessary observation data is not available [78].

Bayesian probability can be interpreted as a numerical degree of belief (between 0 and 1); the measure of the plausibility of an event given incomplete knowledge or the occurrence of other related events. Frequentist views of probability see probability as the long-run expected frequency of occurrence with a belief that the population mean is real. The issue with frequentist views is that random samples for a population may not always be possible or observable. Bayesians believe the population mean is an abstraction with some values more likely to occur than others. This view of probability follows human belief revision and decision making and as such is well designed to represent expert knowledge.

The basic element of Bayesian probability is the *random variable*; a variable whose possible values are outcomes of a random phenomenon. In the absence of any other evidence Bayesian probability uses the *prior* probability. A *prior* probability is the degree of belief that an event will occur given no evidence and is normally obtained from the subjective assessment of an experienced expert. Once agents have observed some evidence which has influence over the previously random variables, *prior* probabilities are no longer appropriate. *Prior* probabilities provide a baseline and as new information becomes available this can be added to the network to develop *posterior* probabilities. In other words, posterior probability is assigned according to given data or evidence while prior probability is not based on any known evidence but on knowledge. *Posterior* probability is based on an event occurring given some observed data. Further to the above classification, probability may either be *conditional* or *unconditional*. *Conditional* probability is where the occurrence of an event is dependent on the occurrence of other events, while *unconditional* probability assumes independence between events. The process of replacing *prior* probability with *posterior* or *conditional* probability is known as *belief network learning* which is the automatic process of determining a suitable belief network, given data in the form of cases. Each case represents an example or situation in the world (presumably that exists or has occurred) and the case supplies values for a set of variables which describe the event.

Bayes theorem is derived from the notion of *conditional* probability. Suppose that *Y* has an influence on the state of *X* then the *conditional* probability of *X* given that *Y* occurs can be written as:

$$P(X|Y) = \frac{P(X \wedge Y)}{P(Y)}$$ and is defined only if P (Y) > 0.

From this formula, Bayes theorem can be derived which is stated as follows:

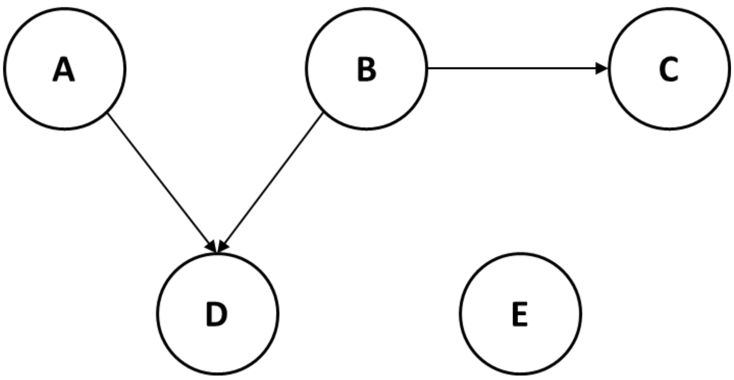
$$P(Y|X) = \frac{P(X|Y)\,P(Y)}{P(X)}$$

Where:
P (Y) = *prior* probability of hypothesis Y
P (X) = *prior* probability of evidence X
P (Y|X) = probability of Y given X
P (X|Y) = probability of X given Y

2.8.3. Formal Definition of Bayesian Networks

A simple BN graph that consists of 5 nodes has been presented in Figure 3. The circles of the graph represent variables and are referred to as nodes. These nodes have a variable name and states. The arrow connections between the nodes are called edges which represent dependence between the variables. Where there is an edge from one node to another, this indicates that one node is the parent of the other and there is a no independence assumption. For nodes where there are no edges pointing to another node, independence is assumed. Per Figure 3 below, there is no edge linking node E to any other node and this implies that E is independent of all other nodes in the graph.

Figure 3: A simple BN graph



Every node of the graph has an associated probability distribution table [80]. In the case of "A", "B" and "E" nodes that do not have a parent node, their probability distribution is an *unconditional* probability. In the case of "C" and "D" which have parents, the probability distribution of these nodes is *conditional* on the parent nodes. The probability of the network is derived as:

P (A, B, C, D, E) = P(A) P(B) P (D |A, B) P (C |B) P(E)

The formal definition for a Bayesian network is expressed as a pair $B = (G, \Phi)$ [80][85]. G is a finite directed acyclic graph whose nodes represent random variables and is represented as $G = (V,E)$ [80][85]. V is a set of vertices (also called *nodes* or *variables*) and E is the set of edges (also called *links*) connected to the vertices [80][85]. For every $x \in V$, G is a finite directed acyclic graph whose nodes represent random variables and is represented as $G = (V, E)$. For every $x \in V$, there is a conditional probability distribution $P_B (x| \pi(x))$ of the node x, given its parents $\pi(x)$ [80][85]. If V is the set of nodes in G, it can be said that B is a Bayesian network over V [80][85].

2.8.4. Decision Networks

A decision network (also known as an influence diagram) is a Bayesian network extended with utility functions and with variables representing decisions. A decision network is solved by computing the strategy that yields the highest expected utility. Influence diagrams are directed acyclic graphs with three types of nodes – chance nodes, decision nodes and utility nodes. Chance nodes, usually shown as ovals, represent random variables describing the environment. Decision nodes, shown as squares, represent the choices available to the decision maker. Decision networks can be viewed as special types of Bayesian network where the value of each decision node is not determined probabilistically but rather is computed to meet some optimization objective [79][80]. A decision network is a very good tool to model the share evaluation process by using probability to represent the domain knowledge [79][80].

2.9. Bayesian Networks in Finance

Bayesian networks may be used in place of ontologies where the decision process is uncertain. Bayesian networks facilitate the implementation of practical AI systems given that they cope particularly well with uncertainty. The uncertainty prevalent in decisions pertaining to investing arise due to limited knowledge about shares and the underlying companies and indeterminism which implies we are unlikely to be able to predict with absolute certainty the outcome of the decision made. Bayesian networks are able to accommodate these challenges.

Not only are ontology-based approaches utilized for automated analysis and decision support; there are several instances of Bayesian networks employed in situations where uncertainty is inherent to the decision process. Cassim [81] presented research on predicting financial distress of JSE listed companies through the use of Bayesian networks and pointed out that the South African data set is somewhat limited for training purposes, having only 66 bankruptcies from 2000 to 2013 off which to create both learning and inference samples; inferring that the limited data set was perhaps the reason behind the low prediction measures in his research. Demirer et al. [77] utilized Bayesian networks to estimate portfolio value, with the assumption that the analyst uses a top-down approach by engaging in economic and industry analysis before proceeding to firm analysis. Demirer et al. [77] provide an illustrative example utilizing a portfolio of Biomedical firms to demonstrate how economic factors (such as, unexpected changes in inflation) beyond firm-specific factors may influence the value of an investment portfolio. Demirer et al. [77] note "a Bayesian network is an important tool to improve security analysis by helping the analyst improve forecasts and eliminate bias". Further to this, they highlight that the most important feature of the Bayesian network is that it allows for dynamically adding new information as evidence as it becomes available. This is particularly important to investment professionals where new information continuously arises and must be rapidly incorporated in their decision making. While the work of Demirer et al. [77] highlighted how a portfolio value may change given changes to certain economic variables, it does not provide a specific share selection framework as laid out in this research.

Bayesian analysis is a relatively objective way to determine the influence of information on the results of the decision-making process in the terms of probability by establishing a relationship of joint and marginal probability. Behavioural finance indicates that humans are susceptible to cognitive shortcomings when trying to assess probabilities [82][83]. Improving probability assessment and decision making has been well researched with results and tools which prove useful [77]. Arguably, the most important feature of the Bayesian network is the ability it provides to add new qualitative or quantitative information as evidence to the model [77]. New evidence can be added through updating probabilities assigned to certain states of a variable in the network or through altering the possible states for a variable.

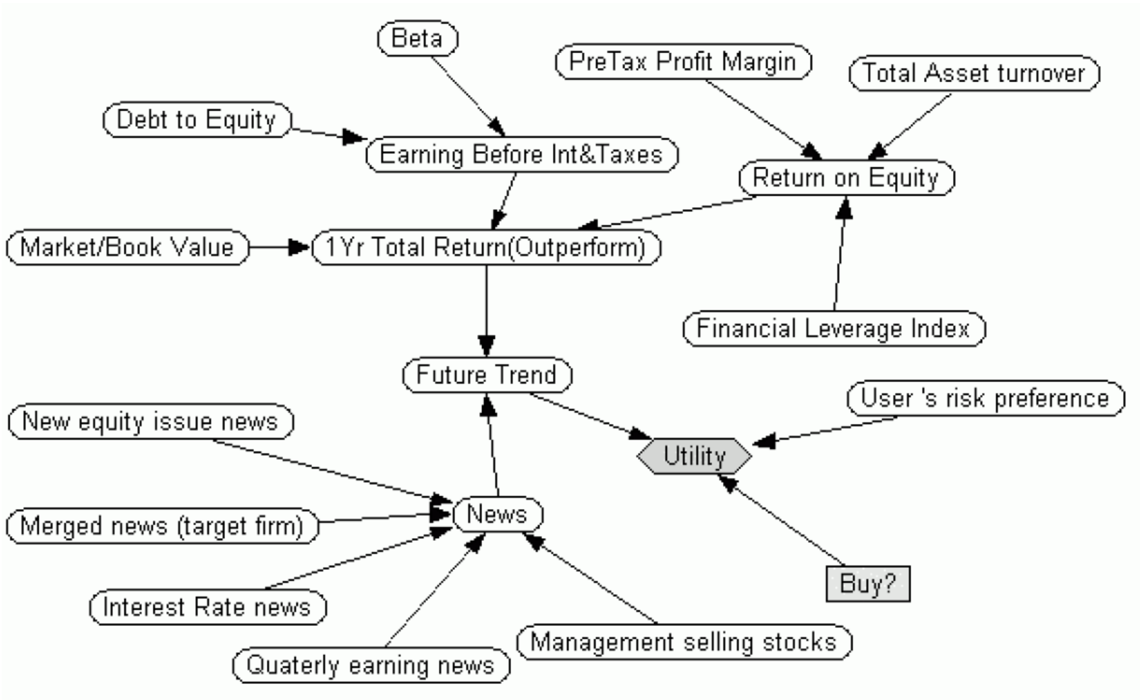
The sub-section below delves deeper into the state of the art with respect to decision networks employed for share evaluation.

2.9.1. Decision Network for Share Evaluation

Two research papers by Tseng [84][79] present a decision network as the decision model for share evaluation and focus on conceptual model refinement. Model refinement is concerned with improving the performance of the model which is measured by the average expected utility of the model run on test cases. While model refinement is important, this work was included to illustrate the process of building a decision network for share evaluation. Tseng's model is by no means complete or clear but provides a starting point to reason with and evaluate. Tseng's network structure was designed through consultation with an expert and utilizes historical financial data to learn probabilities for the variables in the network. Tseng et al. [84] reiterates that the essential issue with a model is to decide on variables and relationships which are important and which can be omitted or are redundant. Tseng et al. [84] highlight that risk is an important factor that varies among portfolio managers. In addition, news about a company's situation may influence the market price trend for a stock. Both factors were incorporated into the model.

Tseng et al. [84] illustrate how the utility in the model depends on the select action (buy or don't buy), risk preference and future trend of the share price. The expert uses subjective values representing the situations and assigns the values to the utility node. Once the initial diagram is constructed with the expert, the next stage was to define the number of values for each variable. While most variables are continuous, they were modeled as discrete. The first prototype was as simple as possible; each value of each variable was carefully defined. For example, the beta value was defined with two states; representing two ranges of beta values. Explicit definitions of the states are necessary to avoid ambiguity when assessing conditional probabilities. Tseng et al. [84] apply an algorithm to refine the number of values attributable to the financial factor nodes like return on equity, debt to equity, beta and so forth. The process for creation of the decision model using an influence diagram was adopted in this research.

Figure 4: Influence diagram for stock portfolio selection from Tseng et al. [84]



2.10. Semantic Bayesian Networks in Finance

Bayesian networks which have been semantically enriched through the creation of an ontology to articulate key concepts to be used in the decision model are known as semantic Bayesian networks. Section 2.7 and 2.9 above have illustrated that ontologies and Bayesian networks serve as useful techniques to represent knowledge in Finance. To our knowledge there are no approaches within the Financial domain that integrate ontologies and Bayesian networks to produce semantic Bayesian networks which could provide automated decision support.

As highlighted earlier, there is a need for more decision support systems to be based on well-defined ontologies. The research papers and models discussed in Section 2.2 and Section 2.9 fail to provide clear definitions and class hierarchies for the inputs and outputs to the decision support systems; making interpretation difficult and ambiguous. Bearing this in mind, our model design makes use of an ontology to provide semantic support to the decision support system which has been modelled using Bayesian networks. **FAST** (discussed in Section 2.7.3) serves as one of the few examples where a decision support system is based on a formalized ontology and this was used as a guide for a new decision model design.

Section 2.2 and Section 2.9 highlight that there are numerous decision modelling tools that may be employed to model the share evaluation process. **FAST** employs SWRL rules to articulate the decision rules and hierarchy. Multi-criteria decision models as reviewed by Aouni et al. [28] and presented in Samaras et al. [30] make use of exact methods which make use of utility functions to weight the evaluation criteria and rank shares accordingly. Tseng [79] utilizes decision networks to articulate the decision process which also make use of utility functions to solve the model. The preferred modelling tool for our approach are decision networks given the ability to deal with uncertainty inherent in the share evaluation decision. Decision networks make use of utility functions to reach an optimal solution and allow the modeler and expert to reason with multiple evaluation factors. The decision network is to be modelled to incorporate a decision hierarchy like that presented in **FAST** which reflects the fact that some evaluation factors are decidedly more important than others and should be evaluated sequentially rather than simultaneously. It follows, that none of the existing models are quite correct in expressing the share evaluation process but a combination of their ideas leads to a more robust decision model. Several researchers who employ Bayesian or decision networks to model decisions have pointed out that the structural information of the decision network is most important to the performance of the system [77]. While Tseng [79] presents a solution for share selection utilizing a decision network, it is unclear how to navigate the decision model with the information provided. A well-defined ontology can alleviate this issue through formalization of the inputs and outputs to the decision process. New model designs should focus on the structure of decision models such that they reflect the manner in which decisions are made in the real world.

2.11. Summary

As noted in this chapter, decision support systems may be created for numerous investment approaches. This research hones in on the value investing approach given that it is the dominant investment approach with widespread adoption. Existing decisions support systems lack clarity of purpose given that they are not explicitly designed for a specific investing approach. Numerous techniques are available for modelling decisions; this research proposes the use of ontologies and Bayesian networks in combination to alleviate the limitations of each. To our knowledge, there is little research on semantic Bayesian networks but this research aims to highlight their importance for the creation of understandable, re-usable and adaptable decision support systems. The next chapter focuses on methodologies for model development. While the design approaches are grounded in ontology development, this does not preclude them from being extended to design semantic Bayesian networks.

3. Methodology for Model Development

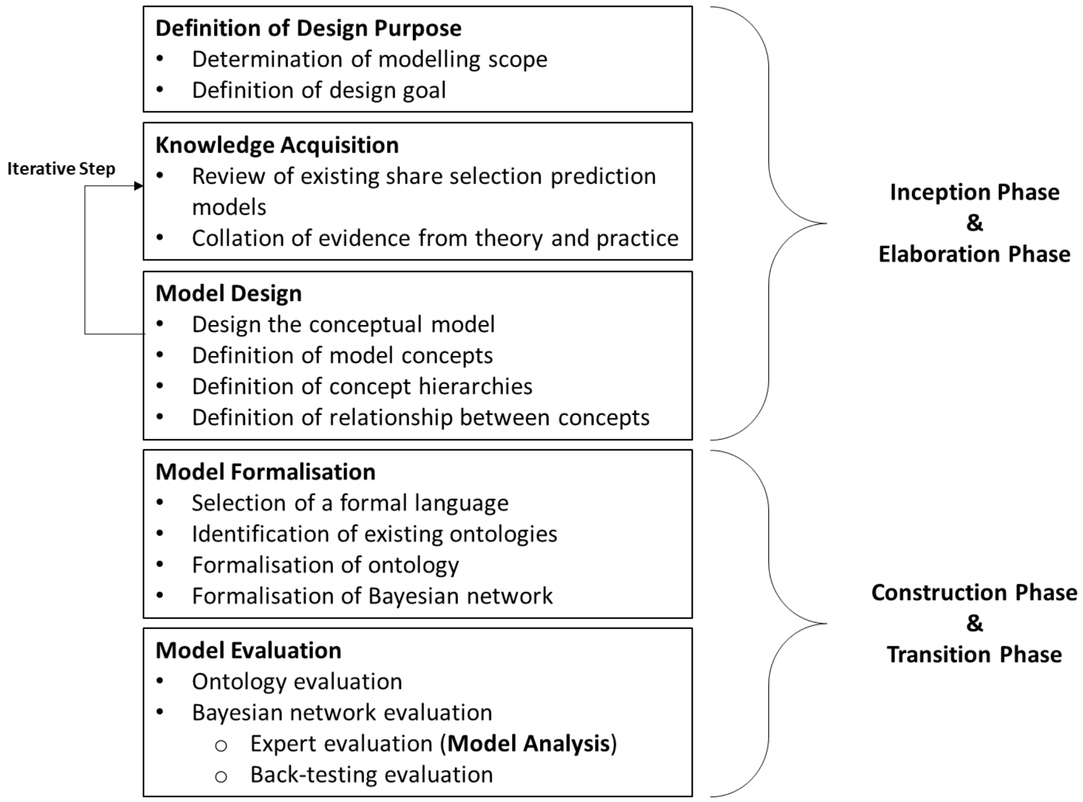
In this chapter, an **ontology-driven approach** as first proposed by Ogundele [85] is described. This approach guides the structuring of knowledge pertaining to the share evaluation process using ontologies and Bayesian networks. As described in Section 2.6, many methodologies exist for ontology development. The UPON methodology was found to be the most suitable candidate for developing an ontology for share evaluation.

3.1. Design Approach

The UPON methodology follows a use-case driven, iterative and incremental approach which makes it unique from other processes [31]. The focus of this approach is to create ontologies to serve the intended users. The iterative nature of the methodology ensures that unclear requirements, prevalent in the design of expert-based systems, are refined over time. The methodology supports the middle-out strategy for identifying concepts and lends itself to the broad domain of share evaluation which contains numerous concepts and complex relationships. Finally, UPON encourages re-use of existing ontologies which alleviates repetition of what is known but also reduces the time and effort required for ontology development.

Ogundele [85] proposes an extension of the UPON methodology to integrate the development of Bayesian networks. This ontology-driven approach has been adopted, where possible, to design, construct and evaluate an ontology and Bayesian network for share evaluation following a value investing approach.

Figure 5: Adapted ontology-driven approach



The first three steps of the adapted ontology-driven approach detailed in Figure 5 are aligned with the activities that are involved in the first two phases of UPON and the last two steps are aligned with the last two phases of UPON. The four phases of the UPON methodology are described in detail in Section 2.6.3.

The adapted ontology-driven approach detailed in Figure 5 above was adapted from Ogundele [85] by incorporating the Model Analysis phase into the Model Evaluation phase. Ogundele [85] provides for an iterative step to revise and refine the initial conceptual model before it is formalized as an ontology. This important iterative step, which is explicitly stated in their approach, provides a feedback loop from the Model Analysis to the Knowledge Acquisition step. The modeler is guided to acquire further knowledge in cases where the model required refinement; allowing for incremental knowledge acquisition and a gradual evolution of the conceptual model. The approach in this research differs from Ogundele [85] given that the iterative feedback step occurs as the conceptual model is designed in the Model Design phase and prior to Model Formalisation.

3.2. Description of the steps involved in the Design Approach

The UPON Methodology has been described in detail in Section 2.6.3. but it is important to reiterate that completion of each phase produces a partially completed version of the ontology. Each cycle results in an improved version of the ontology. Throughout the process the domain expert and knowledge expert worked closely together.

3.2.1. Definition of Design Purpose

The first step is the definition of the goal or purpose for developing the INVEST system. This step involves two activities: (i) determination of the modelling scope and (ii) definition of the design goal and objectives. The original UPON methodology specifies four activities; two of which have been excluded from the process followed in this research as they were specific to ontology development whereas the model design extends to a broader system which encompasses both ontologies and Bayesian networks.

Determination of modelling scope: Since share evaluation is broad and complex, there is a need to constrain the scope of the system design. This is done by constraining the investment approach pertaining to share evaluation.

Definition of design goal: This activity defines the overall design goal for the INVEST system which should conform to the original research question, aims and objectives.

3.2.2. Knowledge Acquisition

Knowledge acquisition is a pivotal, iterative step in ontology engineering. Knowledge acquisition has not been confined to the design chapter but extends to the literature review which provides the body of knowledge necessary to articulate the new conceptual model. The knowledge acquired facilitates the initial identification and definition of terminologies to be included in the ontology. Knowledge to be acquired was twofold: (1) broadly understand the share evaluation process and (2) deeply understand the value investing approach, the evaluation factors and the manner in which these are employed in the decision hierarchy to make an investment decision. The procedures below were carried out to acquire the necessary knowledge. These facilitated the articulation of a conceptual model for the evaluation factors and the design of the share evaluation application.

Extensive literature review: The literature review captured existing research studies in the financial domain which focused on either (1) the broader share evaluation process or (2) the more granular detail about the appropriate evaluation factors for share evaluation under the value investing approach. These research studies articulated the manner in which value investing is translated from theory to practice to reflect how experts carry out share evaluation and what evaluation factors they employ to do so. Following this, state of the art model designs and applications from the computer science domain were evaluated to understand how knowledge from the financial domain on share evaluation is represented and reasoned with and to identify design components which may be re-used, refined or adapted in the new model design.

Consultation with domain experts: While UPON recommends relying on existing research studies in place of domain experts, consultation of domain experts was necessary to assist the modeler in representing the knowledge from existing studies in a model design. Domain experts who conformed to the value investing approach have sufficient community agreement to provide meaningful input on the requirements, concepts and relationships between concepts for the system design.

The knowledge acquisition step is important as it assists the modeler in being familiar with share evaluation domain concepts to evaluate their fitness for inclusion in the conceptual model and system design.

3.2.3. Model Design

The model design uses the knowledge acquired from the prior step to design the conceptual model for representing share evaluation and more specifically, future performance prediction of a share under the value investing approach. A conceptual model is vital for articulating the concrete manner in which share evaluation factors are selected and used to support share evaluation. This step initiates the design process of a concrete model which can be formalized.

Definition of model concepts: Concepts from existing share evaluation models identified through the knowledge acquisition step are modified and restructured. The restructured concepts are to be concretely defined to eliminate overlaps and reduce ambiguity of concepts. Concepts provide a mechanism for consistent knowledge representation; facilitating access, query and navigation of future performance indicators utilized for future performance prediction and selection as carried out in theory and practice.

Definition of concept hierarchies: Some of the factor concepts are hierarchical. This arises from the fact that some are considered as non-negotiable while others are to be weighed against other factors. Non-negotiable factors are those that are a minimum for a share to be considered investable. It follows that the share should meet these minimum hurdles first. These hierarchical relationships should be determined and structured.

Definition of relationships between concepts: Once the concept and concept hierarchy are established, other forms of relationships between these concepts should be defined.

Design of decision model utilizing Bayesian networks: The decision model structure must be designed through consultation with investment professionals on share evaluation. The probabilities are learned from historical financial data.

3.2.4. Model Formalization

The formalization of the conceptual model is done through conversion to an ontology using a formal ontology language. While the ontology is not the only component to the larger system, it serves as the backbone for all other components. The ontology is formalized from a conceptual model by the following three activities:

- **Selection of a formal language** that will be used for the formalization process. The Web Ontology Language (OWL) is the recommended language for formalizing the share evaluation ontology. OWL is the recommended W3C standard that is widely used in several domains, including the financial domain. There are several desktop and web tools for editing, querying, publishing and sharing ontologies that have been formalised with OWL.

- **Identification of existing ontologies** that can be used to build parts of the new ontology will prevent 're-inventing the wheel' in the representation of financial and share evaluation concepts. These ontologies can be existing share evaluation ontologies which can be used as base concepts for building the classes and relationships for the ontology. The existing ontologies can either be directly incorporated or used as a base concept for implementing the ontology classes.

- **Formalisation of the ontology** entails the conversion of the conceptual model into a formalised ontology. The concepts of the model are formalised as classes in the ontology, while the relationships between concepts are represented as properties of the classes. Ontology editing tools are recommended for the construction process. For formalisation in this research paper, Protégé-OWL was used. These tools provide interfaces for easy construction, navigation and querying of the ontology.

Following these three activities, the final step is the **formalisation of the Bayesian network** utilising concepts articulated in the ontology to design the decision model and create a semantically enriched Bayesian network. For formalisation in this research, Netica was used.

3.2.5. Model Analysis & Evaluation

This section deals with the final evaluation of the INVEST system. Most researchers validate their systems with experts or expert research studies but some contend that other validation methods should be used. Colomo-Palacios et al. [76] and Yunusoglu [16] perform back-testing by comparing the performance of the constructed portfolios selected through their decision models to that of a benchmark index. This evaluation method appears more objective than the subjective nature of enlisting experts' opinion in the evaluation. That being said, a combination of evaluation with experts and back-testing can build a more robust evaluation process.

Model Evaluation: Ontology
The ontology was evaluated to validate its usefulness for the purpose for which it was created. This was done by outlining usage examples and identifying how the ontology serves the broader INVEST system.

Model Analysis with Experts: Bayesian network
An initial version of the model is to be analysed with experts to ensure its comprehensiveness and also to identify aspects of the model that need improvement prior to the back-testing evaluation. The inputs to the model, the model hierarchy and decision outputs are to be evaluated by experts. Following this, the model is refined and subjected to further evaluation through back-testing.

Model Evaluation through Back-testing: Bayesian network
The final evaluation was carried out using historical data to evaluate the performance of the designed decision model.

3.3. Summary

The ontology-driven design approach which is an extension of the UPON methodology was adopted. This chapter details the steps involved to carry out the design approach with reference to the research problem at hand. The design approach is applicable to both the design of the ontology and Bayesian network and more broadly to the INVEST system to be designed.

4. Model Development and System Design

This chapter describes the application of the ontology-driven approach, as presented in Chapter Three, and the resulting system which encompasses the ontology and Bayesian network.

4.1. System Goal & Objectives

Share evaluation spans multiple investment approaches. Each investment approach reflects an investor's belief of what drives future share performance and requires careful study and expert consultation in order for the relevant knowledge to be made explicit through an ontology. It is important to have well defined limitations on what the system is expected to know and what its capabilities should be. While share evaluation is the overarching process followed by any investment professional; each investment approach should be considered a new domain. The more domains that are added to the system, the more complex the system becomes and as such the problem should be limited to a sufficiently narrow scope [86]. Bearing this in mind, the system shall not be exhaustive with respect to share evaluation; the decisions and knowledge in this arena are complex and infinite. The system provides a share evaluation framework for extension for any investment approach, but the implementation of the system has been constrained to the value investing approach. Where possible, key concepts applicable to share evaluation and common to all investment approaches have been articulated. To implement the system design, the value investing approach has been adopted to illustrate the application of the system to a concrete problem. Thus, the lower level concepts and concrete classes are specific to a value investing approach. Data pertaining to shares listed on the JSE has been used for validation of the Bayesian network.

The main goal of the INVEST system is to capture and reason with explicit knowledge pertinent to share evaluation utilizing ontologies and Bayesian networks. The novelty of this research is the implementation of the system design through a concrete application of the value investing approach which is used to create an ontology to articulate key investment concepts which are then utilized as inputs to the Bayesian network.

This goal includes the following **objectives**:

(1) Design an ontology which captures domain knowledge for share evaluation under the value investing approach
(2) Design a Bayesian network to capture the complex decision process for share evaluation under a value investing approach
(3) Design a system which supports the ontology and Bayesian network and facilitates user query

The next section articulates how knowledge was acquired in order to design the above components of the INVEST system.

4.2. Knowledge Acquisition

The share evaluation process is rarely described or documented in isolation in literature but is implicitly expressed through documentation on specific investment approaches. The value investing approach was selected to elicit an understanding of the share evaluation process and more specifically understand the manner in which an investor selects and categorizes share evaluation factors. Section 2.3. provides an overview of value investing in theory and practice, and Section 2.7. and Section 2.9. review existing share evaluation models. These serve to provide the background knowledge required to understand the share evaluation process and how it is applied in a specific context. The construction of effective predictive models is challenging and relies on detailed expert knowledge to identify factors which influence future share performance. While literature on investment theory is abundant, there is no known concrete, unambiguous and comprehensive computer based conceptual model that actively categorizes factors that are indicative of or influential to future share performance. The reference lexicon which details the factor categorization model below attempts to address this gap.

4.2.1. Reference Lexicon

A reference lexicon was developed for share evaluation and is informed by Section 2.3. to address the challenges of existing share evaluation models as highlighted in Section 2.7. and Section 2.9. This serves as a guiding reference for the categorization model used for the key concept to the ontology: **Factor.**

Table 12: Key concepts for the Share Evaluation Domain

Main Terminology	Description
Factor	A factor is any ratio, figure or qualitative variable that is believed to be indicative or influential to future share performance.
Factor Type	Categorization of evaluation factors by theme or type involves the grouping of factors based on the main subject or topic that defines their similarities. For instance, all factors which are return metrics may be grouped together. Themes like profitability, growth and risk have been developed from the review of existing categorization models.
Evaluation objective / question	Categorization of evaluation factors based on the evaluation objective. For instance, value investors objective is to determine whether the *price is reasonable* and such factors which serve to answer this should be grouped accordingly in a 'Value' group. The focus shifts from the type of factor to understanding what question the factor may answer for the investor. A specific investment approach will dictate the respective evaluation objectives. An investment approach may have more than one evaluation objective. For instance, value investors wish to determine whether a share *is reasonably priced* (determine **Value**) and *is of a high quality* (determine **Quality**).

4.2.2. Conceptual Model Design

The conceptual model is limited to the factors required for share evaluation. Some of concepts relevant to the share evaluation process, such as Investment Asset and Classifier, are excluded from this step but presented in the formalized model in Section 4.4. These concepts have been adequately evaluated and validated in prior research studies.

4.2.2.1. Existing Factor Categorization Models

The Literature Review illustrated that decision model objectives for share evaluation are vague and unclear; it is thus unsurprising that share evaluation factors are poorly categorized. The discussion below draws on existing decision models discussed in the Literature Review to illustrate existing factor categorization models as summarised in Table 13.

Table 13: Summary of Existing Factor Categorization Models presented in literature

Dimension	Colomo-Palacios et al. (2011) – FAST Ontology	Aouni et al. (2018) – Review of MCDSS	Samaras et al. (2006) - MCDSS for Share Evaluation	Tseng (2006), Tseng et al. (2001) – Decision Network	Yunusoglu (2012) – Fuzzy-based expert system
Factor Type		**Financial criteria** which includes profitability, liquidity and operating performance ratios	**Investor preference criteria**	**Financial criteria**	Elimination factors based on **investor preference criteria**
		Non-financial criteria			
		Stock market criteria which focus on valuation ratios like price-to-earnings			
		Investor preference criteria			
		Risk criteria which include variance, skewness, kurtosis etc.		**Risk criteria**	
		Return criteria		**Return criteria**	
Period of Prediction	**Long-term prediction criteria**				
	Medium-term prediction criteria				
Evaluation Objective	**Value criteria**				**Fundamental criteria**
	"Good Company to Invest in" criteria				**Technical criteria**
Industry/ Accounting Grouping			Commerce/Industry		
			Insurance		
			Investment		
			Bank		

The FAST ontology names and defines the share evaluation factors but has not correctly categorized them; confusing factors that indicate whether a stock is cheap or expensive with those that indicate that the share represents a high-quality company. The existing works presented by Aouni et al. [28] and Samaras et al. [30] highlight the continued expansion of share evaluation factors which may be incorporated into a decision model. Deliberation over the factors incorporated in the decision model is not prevalent in their work. It appears that the manner in which these decision models have been formed lack clear categorization and procedural methods for identifying key share evaluation factors. This presents a challenge for a modeler with little to no knowledge of share evaluation and leads to poor selection of the correct share evaluation factors to satisfy the decision model objective. This challenge is to be addressed in this research through a new conceptual model which defines a clear approach for evaluation factor categorization.

It was noted that most multi-criteria decision models categorise criteria as specific to a certain sector, industry or accounting plan. On the other end, FAST provides a one-size fits all approach to fundamental analysis which may be inappropriate for the analysis of certain firms. Certain ratios like the price-to-book ratio are only appropriate in the analysis of firms from specific industries. Neither of the above approaches is desirable as a portfolio manager not only wants to perform industry-specific analysis but broader market analysis across shares. A portfolio manager will often evaluate a share on multiple levels: relative to the market, relative to the sector, relative to peers within the sector and relative to its own share price and history. This means that both universal fundamental analysis and industry-specific fundamental analysis should be possible. An ontological approach to support decision modelling would enforce the necessary identification and deliberation of common share evaluation factors, with the ability to extend the model to include industry-specific factors. This research aims to identify universal share evaluation factors and leaves the identification of industry-specific factors to future work.

Most models present means by which the model can be adapted for investor preference and choice. While all value investors may agree that certain share evaluation factors should be considered, the thresholds for these factors should be variable. Samaras et al. [30] allows an investor to specify their *Lowest Stock Profile* as the cut-off to identify shares which are investable. In addition, the investor is able to choose the level of risk as a further factor for share evaluation. Tseng [79] follows suit with a "user's risk preference" incorporated into their influence diagram. FAST allows for a margin of safety to be chosen by the investor. Given that uncertainty exists with most calculated values for share evaluation factors, it is important to leave a 'margin of safety' when making investment decisions. Incorporating variable thresholds for share evaluation factors and the related rules not only incorporates a margin of safety but ensures that the model is easily adaptable to investor preference and choice and should be incorporated where possible.

Giving consideration to existing categorization models; a proposed factor categorization model is detailed below.

4.2.2.2. Proposed Factor Categorization Model

The categorization dimensions identified from the literature review and summarized in **Table 13** were restructured to form components that were used to design the new conceptual model. These have been restructured to present a unique and useful categorization for factors used to evaluate shares. Four dimensions were identified from the review of existing categorizations; these are: *Factor Type, Period of Prediction, Evaluation Objective* and *Industry Grouping.* The literature review, specifically section 2.3.1. identifies value investing share evaluation factors and highlights that an effective categorization model is based on the evaluation objective guided by the investment approach. For this reason, ***Evaluation Objective*** and ***Factor Type*** groupings have been included in the conceptual model design. *Industry Grouping* and *Period of Prediction* are outside the scope of the current model design but should not be considered irrelevant.

The classifications found in the existing studies were used to develop unique and specific **Factor Type** categories. The existing categories were restructured to eliminate concept overlaps and misrepresented factors. They were iteratively checked in terms of their effectiveness to classify factors found in research publications. The process of restructuring categories involves matching existing categories based on similarity of names and meaning. Similar **Evaluation Objective** categories were merged to produce a comprehensive category. Broad categories that represent heterogenous factors were split to produce unique categories without unnecessary overlap. Through the process, four **Evaluation Objective** categories were defined and their boundaries were set to facilitate the inclusion of factors from value investing theory and practice. They are ***Value, Quality, Elimination*** **and** ***Preference***. The process also highlighted seven **Factor Type** categories as follows: ***Present Discounted Value*****,** ***Valuation Multiple, Relative Ratios, Profitability Ratios, Growth Ratios, Financial Risk Ratios*** **and** ***Systematic Risk Ratios***.

A hierarchical model was introduced to capture the two types of categorizations employed in the model in a consistent manner. The top level of the hierarchy includes the categorization termed **Evaluation Objective** while the second level represents **Factor Type** categorizations which fall within these objectives. The lowest level in the hierarchy represents concrete and measurable factors.

The proposed factor categorization model is utilized in the INVEST Ontology in Section 4.4.1.1. and presented through the **Factor** class hierarchy. Prior to delving into detail on the INVEST Ontology, an overview of the INVEST System is necessary to provide context for the individual components to the system.

4.3. The INVEST System

This section serves to describe the design of the proposed system, i.e. the INVEST system, for automatic share evaluation.

4.3.1. Motivation for Design

The INVEST system has been designed to support investment professionals with their medium-term share evaluation decisions in a flexible, practical and realistic manner. The INVEST system has been described in its entirety but the key focus of this work is on the ontology and Bayesian network used to represent and reason with expert knowledge to make share evaluation decisions. The decisions as implemented in this system involve a combination of factual and heuristic knowledge. The system has been designed such that each component is responsible for either a data, knowledge or control structure such that a user may easily retrieve the knowledge contained within the system [87]. The system is able to combine facts and heuristics and thus merge human knowledge with computer power to solve problems [87]. The knowledge and required reasoning have been effectively matched to the right tool within the INVEST system. While the underlying components are separate, the INVEST system has been designed to support user query; it pulls together the information inputs and outputs across the components. An overview of the system is provided below.

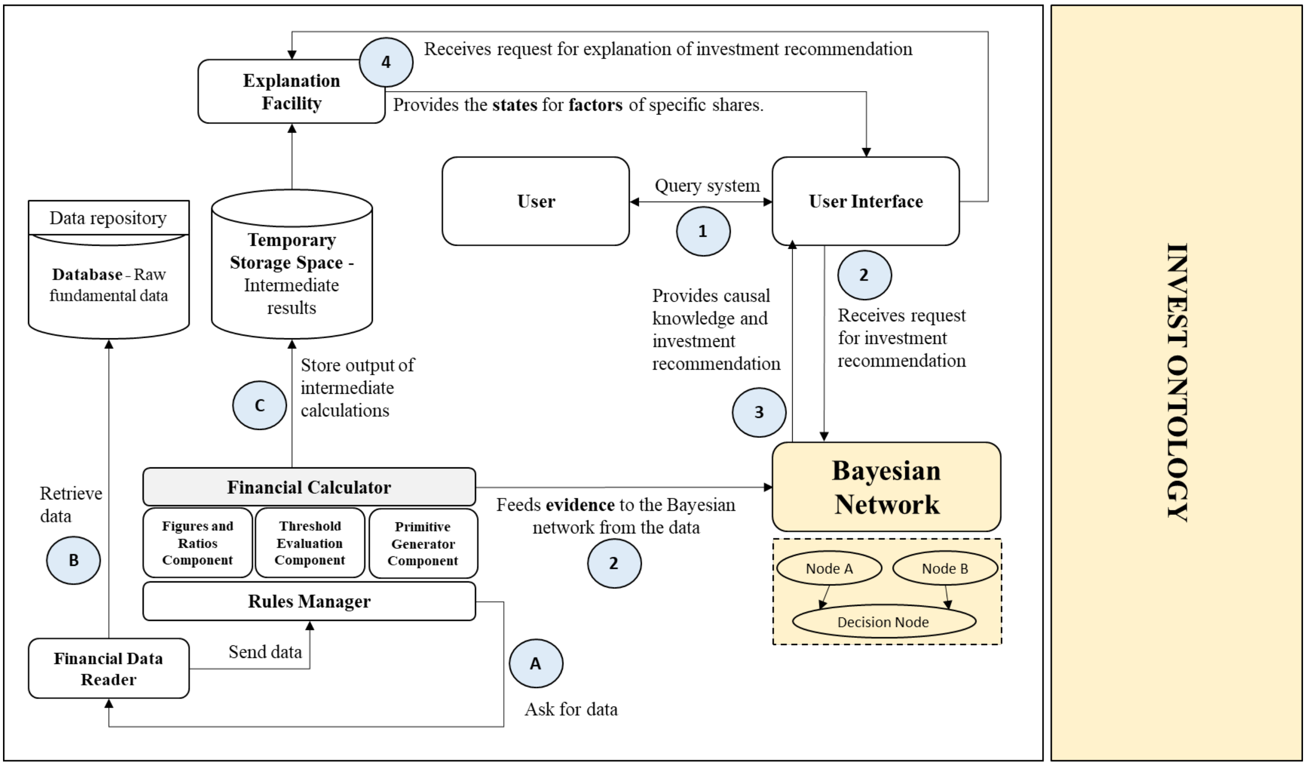
4.3.2. System Overview

The system design has been illustrated below in Figure 6. The design comprises distinct interactive components which are described below.

The **Database** stores raw fundamental data and is updated through a link to a live data feed. The **Temporary Storage Space (TSS)** stores the output of intermediate calculations which are executed by the **Financial Calculator (FC)**. The **Financial Calculator (FC)** requests data from the **Database** and performs calculations utilizing the data through a conventional programming language like Python. The **FC** is composed of three components: **Calculator Component (CC), Threshold Evaluation**

Component (TEC) and **Primitive Generator Component (PGC)**. The **INVEST ontology (IO)** represents expert knowledge on share evaluation through concepts, properties and relationships between concepts. The **IO** captures concepts across the entire **INVEST system.** The **Bayesian Network (BN)** represents heuristic expert knowledge for share evaluation which is composed of intuition, judgement and logical inference through causal knowledge stipulated using a Bayesian network. The **BN** evaluates the evidence provided from the **FC** using causal knowledge to reach intermediate conclusions and a final investment recommendation. The **Explanation Facility (EF)** provides an explanation to the user as to how a certain decision was reached by the **BN**. The **EF** component presents the rules used to derive the specific recommendations.

Figure 6: The basic structure of the INVEST System



4.3.3. System Workflow

The workflow of the system is described below. The steps detailed below correspond to the numbers in Figure 6 above.

1) The **user** is able to query the system through a user-friendly **interface** to request investment recommendations for a certain share. For example, is the share investable or not?
2) Once the query has been received, the **BN** utilizes the encoded causal knowledge to evaluate the evidence for a particular share provided by the **FC** to provide an investment recommendation. There is a subset of processes which ensue to produce the evidence contained in the **TSS**.
   A. The evidence contained in the **TSS** are the result of calculations performed on raw fundamental data stored in the **Database**. When evidence is requested from the **FC** by the **BN**, this triggers the **rules manager** contained within the **FC** to ask the financial data reader for the necessary data to execute each calculation rule to produce the evidence for the **BN**.

B. The **financial data reader** module queries the **Database** to retrieve the information and return back to the **rules manager**. Once the information has arrived at the rule manager, it will make interchange calls with the **FC** module that is in charge of executing necessary financial operations.
C. The **FC** has a **TSS** set aside for storing the results of intermediate calculations.

3) The **BN** follows the decision hierarchy to reach intermediate decisions and a final recommendation through passing the evidence through the encoded causal knowledge. The final recommendation result is returned to the user in an investment answer format.
4) The **EF** provides a mechanism by which the user can query the reason for the investment recommendation. The **EF** facilitates the transfer of the **states** of the **factors** used to reach an investment recommendation to the user interface.

The overview of the INVEST system above is followed by details on the design and selective implementation of key components in the INVEST system which focus on knowledge representation. The value investing approach serves as the case study to guide the development and implementation of a prototype of the INVEST system and reflects how the system can be used in real world application.

4.4. The INVEST Ontology

The INVEST ontology (IO) provides a clear structure for information that may be useful to an investment professional and articulates concepts and properties which are required as evidence for the Bayesian network**.** The ontology was developed using Protégé-OWL tool version 5.5.0 to implement the ontology in OWL, including design of the ontology and populating the ontology with instance data.

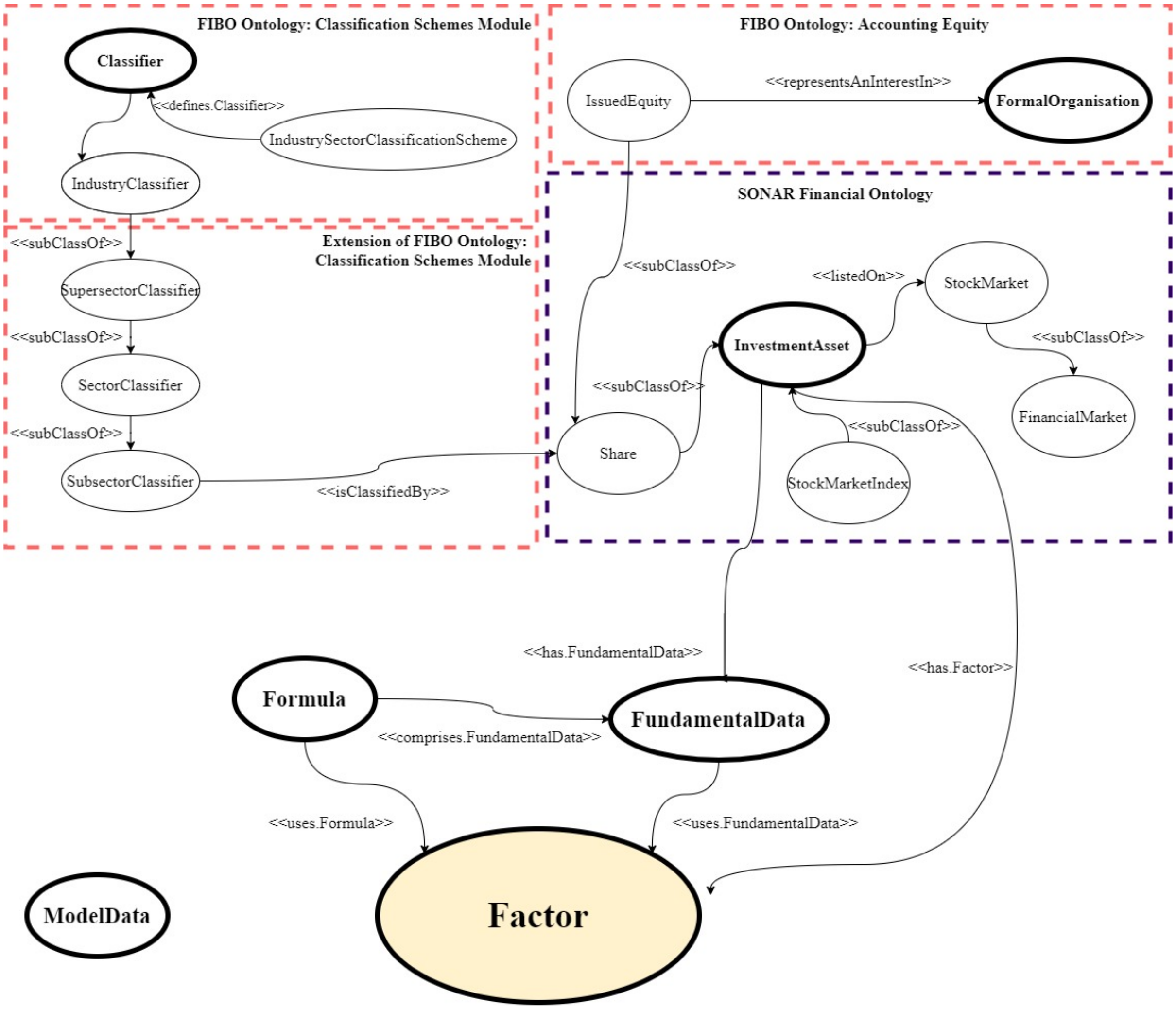
Concepts from the following ontologies were reused:

- Financial Industry Business Ontology ('FIBO') Foundations Version 1.2 [88] for the **classifier** concept and **issued equity** concept pertaining to a formal organisation.
- The SONAR financial ontology [89] for the **investment asset** concept.

4.4.1. Main Classes and Properties

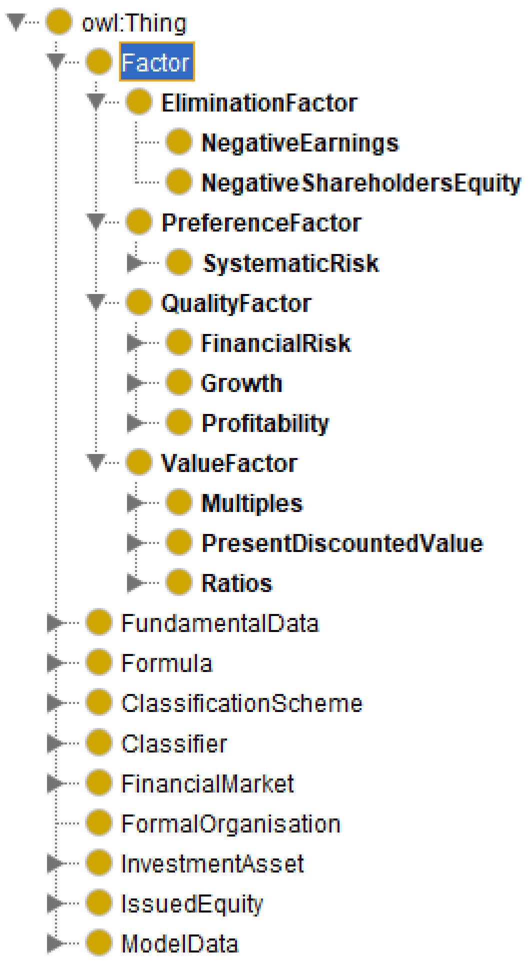
The main classes of the ontology are common to any share evaluation process; they are not confined to the value investing approach. The ontology consists of eight main classes: **Factor; Formula; FundamentalData; ModelData; InvestmentAsset; FormalOrganisation** and **Classifier**. These determine the structure of the database for the INVEST system.

Figure 7: Overview of the key concepts, properties and relationships of the ontology



The key classes, properties and relations of the ontology are captured in Figure 7. Figure 8 below shows classes of the ontology in the Protégé ontology editor. See Appendix 1 for the complete OWL representation of the INVEST Ontology.

Figure 8: Classes of the INVEST ontology as represented in Protégé ontology editor



4.4.1.1. Factor

The **Factor** class is any ratio, figure or qualitative variable that is believed to be predictive or influential to future share performance. The factors identified and implemented are informed by research studies and the guidance of experts pertaining to the value investing approach. The **Factor** class is constructed hierarchically with sub-categories of factors represented by three sub-classes. The first level of the class hierarchy is the abstract **Factor** class and the second level includes the abstract classes: **ValueFactor**, **QualityFactor, PreferenceFactor** and **EliminationFactor**. The lower levels of the class hierarchy represent concrete observable factors. For example, **ForwardPE_CurrentvsHistory** falls under the **ValuationMultiples** sub-category and **ValueFactors** main category (see Figure 9 below). The hierarchical classes are used to represent factors as categories and sub-categories. The **evaluation objective** categorization developed in the conceptual model (see Section 4.2.2.) is implemented for the main class of the **Factor** class hierarchy. Following this, the **factor type** categorization developed in the conceptual model (see Section 4.2.2.) is used to classify sub-categories of factors in the class hierarchy.

It is important to note that a specific factor may belong to multiple evaluation objective categories; this allows for ambiguity and the different perceptions on which evaluation objective a factor addresses. Although the categories were formulated to eliminate concept overlaps and misrepresented factors, there is still a possibility of a factor belonging to more than one category. Formalizing the model with an ontology allows for multiple association of factors with more than one class. The hierarchical representation of the **Factor** is shown in Table 14 below. The list of factor categories is not exhaustive but details factor categories specific to value investing. Literature and empirical studies would suggest there are far more factor categories that could be included. However, the intention of the model is to provide a framework for extension and to illustrate the application to the value investing approach for share evaluation.

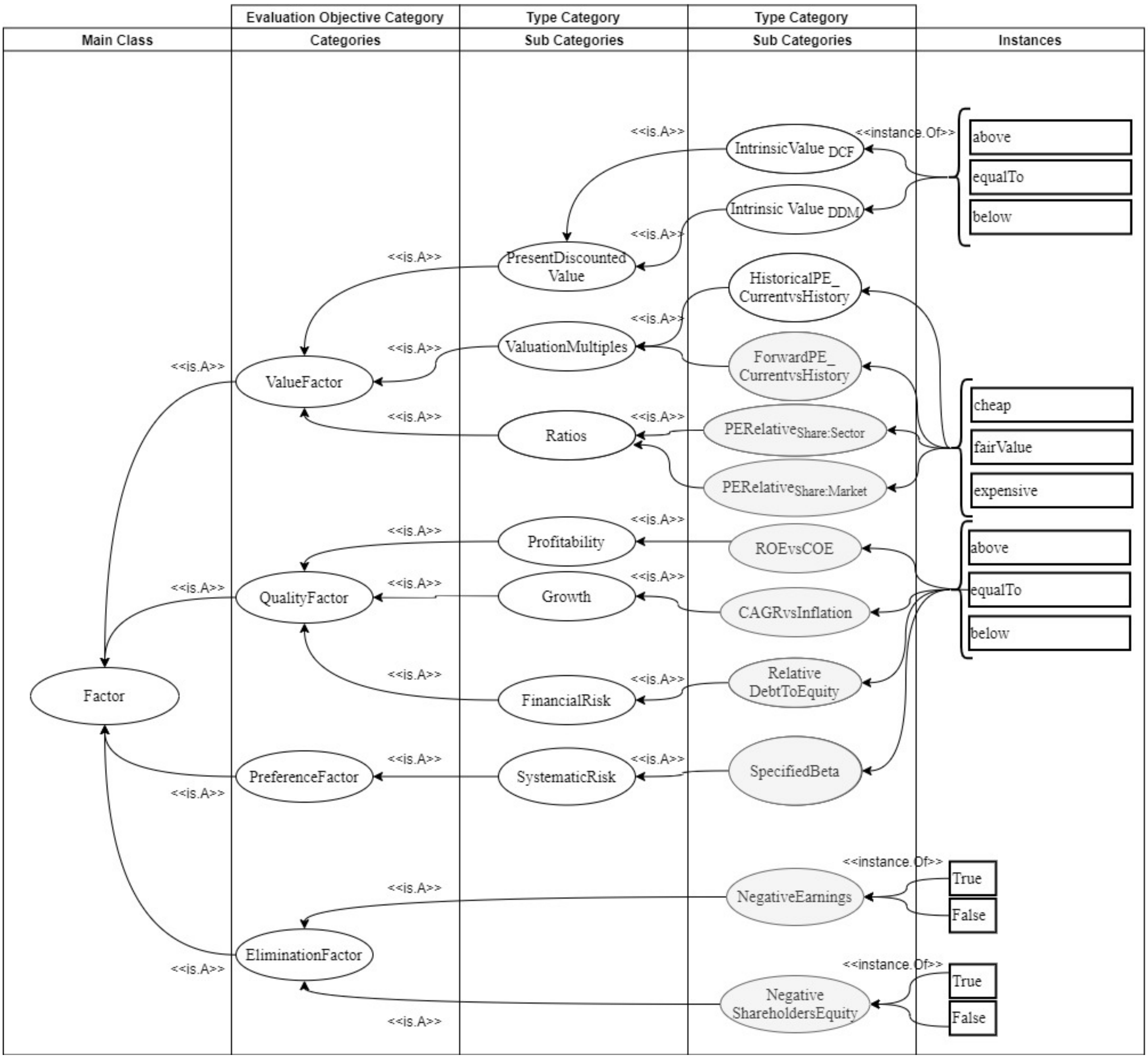
In representing the **Factor** class, a design decision was made to represent factors as the main subject of the model and not as characteristics of a share. This decision is to facilitate structuring of the factors for construction of a predictive model. This is also in line with the approaches taken by all existing categorizations (see [76]and [16]), which have been proposed by domain experts in this area.

Table 14: Sub-classes of the Factor class

Main Class	**Middle Class**	**Bottom Class**
Value Factor	Present Discounted Value	Intrinsic Value $_{DCF}$
		Intrinsic Value $_{DDM}$
	Valuation Multiples	HistoricalPE_CurrentvsHistory
		ForwardPE_CurrentvsHistory
	Ratios	PE Relative$_{Share:Market}$
		PE Relative$_{Share:Sector}$
Quality Factor	Profitability	ROEvsCOE
	Growth	CAGRvsInflation
	Financial Risk	Relative Debt to Equity
Preference Factor	Systematic Risk	SpecifiedBeta
Elimination Factor	N/A	Negative Earnings
	N/A	Negative Shareholders Equity

Figure 9 shows a conceptual diagram of the **Factor** class. It is important to note that the Conceptual Model Design (see section 4.2.2.) is encapsulated in the Factor class which incorporates the Evaluation Objective and Factor Type as sub-classes (see further detail below in 4.4.1.1. Factor). Figure 9 illustrates the class hierarchy and groupings and the types of instances which belong to the concrete classes of the Factor class. For example, the PE relative$_{Share:Market}$ has three discrete instances which indicate whether a share is “cheap”, “fairValue” or “expensive”. These discrete instances have been created through the evaluation of the current value of the PE relative$_{Share:Market}$ against a threshold; for this particular factor it would be the historical PE relative$_{Share:Market}$. Each concrete factor class will have a different set of instances depending on the threshold against which it is evaluated. The concrete classes highlighted in grey correspond to the variables used in the Bayesian networks.

Figure 9: Example of hierarchical class of the INVEST Ontology



4.4.1.2. FundamentalData

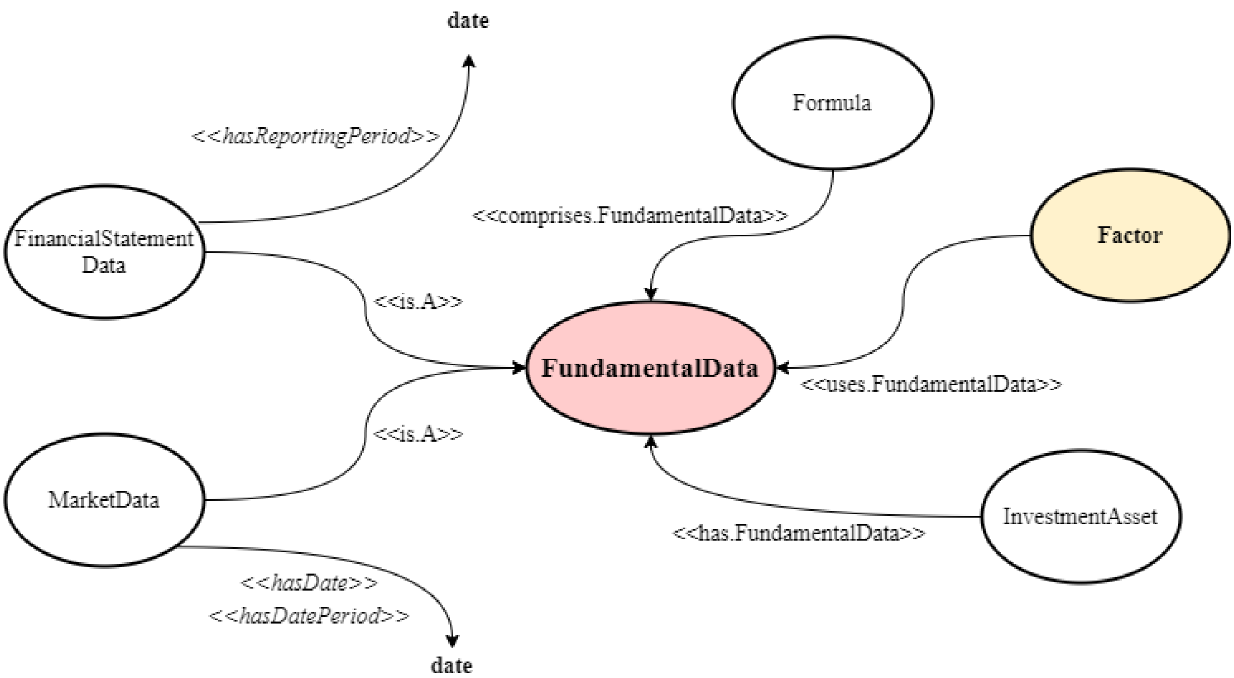
Fundamental data represents raw data that relates to a specific share and is required for share evaluation through fundamental analysis. The **FundamentalData** class has been broken down into further classes: the **FinancialStatementData** class and **MarketData** class. Fundamental data is stored in the **Database** and is updated through a data feed. Instances of the **FundamentalData** class are generally comprised of continuous data values. Fundamental data is used by the **Formula** class, and in some cases by the **Factor** class where no further processing of the data is required.

Table 15: Sub-classes of the FundamentalData class

Main Class	Middle Class	Bottom Class
Financial Statement Data	Income Statement Data	Net Income Shares in issue Earnings per share
	Balance Sheet Data	Total assets Total liabilities Total equity Cash and cash equivalents Total Shareholder's Equity
	Cash Flow Statement Data	Cash from operating activities Cash from financing activities
	Equity Statement Data	Retained Earnings
Market Data	Time series data	Price
		Beta
	Consensus forecasts	Forward Earnings

Data properties are associated with the **FundamentalData** class for validation of the fundamental data to be stored in the database. Where applicable, **hasDate**, **hasDatePeriod** or **hasReportingPeriod** define the date or the period within which the fundamental data is applicable. The relationship of **InvestmentAsset** class to the **FundamentalData** class is classified through the **has** relationship. Depending on the **InvestmentAsset**, there may be certain classes or instances of fundamental data which do not exist for it.

Figure 10: Extract of FundamentalData class, related properties and relationships



4.4.1.3. ModelData

ModelData represents inputs which are applicable to all shares and are used within the decision model. The **VariableParameters** class is used to represent the variable inputs and assumptions of the investor and can be manually updated to reflect the investors preference or beliefs. For example, the investor may have a certain margin of safety, expressed as a percentage, that they may want to apply when evaluating any values in the model. This could be instituted utilizing the **MarginOfSafety** class stipulated below. The **EconomicData** class is updated as new economic figures are published.

Table 16: Sub-classes of the ModelData class

Main Class	Middle Class	Bottom Class
EconomicData	-	Inflation
	-	RiskFreeRateOfReturn
VariableParameters	Threshold	MarginOfSafety
		BetaThreshold
		MarketRateOfReturn

4.4.1.4. Formula

The fundamental data as represented by **FundamentalData** class is processed further through calculation rules stored in the **Financial Calculator** to produce calculated figures and ratios as represented by the **Formula** class. The calculation rules for these figures and ratios are specified in **Section 4.5.1.** The properties described in **FundamentalData** class are replicated for the **Formula** class.

Table 17: Sub-classes of the Formula class

Main Class	Middle Class	Bottom Class
Formula	Calculation formula	GrowthRate
		HistoricEarnings_GrowthRate
		HistoricEarnings_CAGR
		ForwardEarnings
		ForwardEarnings_CAGR
		HistoricPriceToEarnings
		ForwardPriceToEarnings
		PErelative
		Debt/Equity
		RelativeDebtToEquity
		EquityRiskPremium
		ReturnOnEquity
		CostOfEquity

An intermediate rule set is required to evaluate the calculated figures and ratios as represented by the **Formula** class against either a threshold as represented by the **Threshold** class which is a sub-class of the **ModelData** class or against another figure or ratio to produce discrete states which represent instances of the **Factor** classes. This intermediate calculation set, named the **Threshold Evaluation Component (TEC)**, is contained within the **Financial Calculator** component (see Section 4.5.2.). For example, **ForwardPriceToEarnings** may be evaluated against the average **HistoricalPriceToEarnings** for a specific share utilizing the rule set contained in **TEC**. Following this, the instances that may be observed for **ForwardPriceToEarnings** are “cheap”, “fairValue” or “expensive”. This design decision will facilitate the conversion of factors and related instances into useful primitives which can be mapped to Bayesian network variables (see Section 4.5.3.).

4.4.1.5. Classifier

A **Classifier** is a standardized classification or delineation for something, per some scheme for such delineation, within a specified context. The **IndustrySectorClassificationScheme** represents a system used for allocating classifiers to organizations by industry sector. The Financial Industry Business Ontology (FIBO) was adopted as the base ontology to describe the **Classifier** class and related **IndustrySectorClassificationScheme** class. FIBO was extended to include the **SuperSectorClassifier** class, **SectorClassifier** class and **SubsectorClassifier** class which represent standard classifications under the Global Industry Classification Standard (GICS) [90]. The GICS is the classification scheme used for share classification on the Johannesburg Stock Exchange. The **Share** class is classified by the hierarchy of **Classifier** classes.

Table 18: Sub-classes of the Classifier class and Arrangement class

FIBO Class	Sub Class	Sub Class	Sub Class	Bottom Class
Classifier	Industry Sector Classifier	Super Sector Classifier	Sector Classifier	Subsector Classifier
Arrangement	Classification Scheme	-	-	Industry Sector Classification Scheme

4.4.1.6. FormalOrganisation

The **Corporation** class is an extension of the **FormalOrganisation** class which is laid out in the base ontology, FIBO, and was included in the INVEST ontology to reflect the fact that a share represents an interest in an organisation. This is an important concept as an organisation may issue more than one type of share which forms part of issued equity.

Table 19: Sub-class of the FormalOrganisation class

FIBO Class	Sub Class
FormalOrganisation	Corporation

4.4.1.7. InvestmentAsset

An **InvestmentAsset** represents everything of value which a financial intermediary can invest in, such as stock market index funds and shares. The financial ontology as developed in SONAR was adopted as the base ontology for describing the abstract **InvestmentAsset** class and more specifically the concrete **Share** class. This class was included to incorporate pertinent information required for share identification and classification.

Table 20: Sub-class of the InvestmentAsset class

Main Class	Bottom Class
InvestmentAsset	Share
	Stock Market Index Fund

The following data properties are associated with the **InvestmentAsset** class: **hasName** and **hasCode** which capture the name and code for the specific investment asset. For certain classes such as **Share** and **StockMarketIndexFund** that may be listed further properties are defined like **listedOn** which describes the stock market the investment asset is listed on.

4.5. Financial Calculator

The financial calculator (FC) is used to represent the parts of the system that deal with computational or deterministic problems. Conventional programming may be used to implement the calculations and rules specified in the financial calculator. Pseudo-code is used to illustrate a prototype implementation of the system design. The definitions for the concepts used below have been specified in the INVEST Ontology.

There are three distinct sub-components within the Financial Calculator as follows:

(1) Calculator Component
(2) Threshold Evaluation Component
(3) Primitive Generator Component

4.5.1. Calculator Component (CC)

This component is required to transform fundamental data into figures and ratios.

Calculation Rule 1: Historic Earnings Growth Rate

$$\boldsymbol{GrowthRate}_{\boldsymbol{Year\,x+1}} = \frac{\mathrm{EPS}_{\mathrm{Year}\,x+1}}{\mathrm{EPS}_{\mathrm{Year}\,x}} \qquad \text{(Calculation 1)}$$

$$\boldsymbol{HistoricEarnings\_GrowthRate} = \frac{(GR_{Year\,x} + GR_{Year\,x+1} + \ldots + GR_{YearN})}{N} \qquad \text{(Calculation 2)}$$

Where:
EPS = Earnings per share
GR = Growth Rate

Calculation Rule 2: Historic Earnings Compound Annual Growth Rate

The use of historical growth rates is one of the simplest methods of estimating future growth in earnings. However, high historical growth rates do not always indicate a high forward-looking growth rate as industrial and economic conditions change constantly and are often cyclical. That being said, when there is no alternative to historical growth rates such as forward-looking consensus earnings forecasts this is the best measure an investor may have. It is generally recommended that the growth rate is calculated over 5 years or shorter.

$$\boldsymbol{HistoricEarnings\_CAGR} = \left(\frac{EPS_{Year\ N}}{EPS_{Year\ N-x}}\right)^{1/x} -1 \quad \text{(Calculation 3)}$$

Where:
x = number of years
N = final year
EPS = Earnings per share

Calculation Rule 3: Historic Price to Earnings

$$\boldsymbol{HistoricPriceToEarnings} = \frac{Price_{Share}}{Earning\ per\ share} \quad \text{(Calculation 4)}$$

Calculation Rule 4: Forward Earnings

Forward earnings may not need to be calculated where forward-looking consensus earnings forecasts are available. These are often available on Bloomberg and other stock market data sources and have been inputted from human experts who have applied their mind to the growth of future earnings. Where no consensus forecasts are available, it is necessary to calculate a forward earnings figure as per the formula below:

$$\boldsymbol{ForwardEarnings} = EPS_{\text{Latest Financial Year End}} \times HistoricalEarnings\_GrowthRate \quad \text{(Calculation 5)}$$

Calculation Rule 5: Forward Earnings Compound Annual Growth Rate

$$\boldsymbol{ForwardEarnings\_CAGR} = \left(\frac{ForwardEarning_{Year\ N}}{ForwardEarnings_{Year\ N-x}}\right)^{1/x} -1 \quad \text{(Calculation 6)}$$

Calculation Rule 6: Forward Price to Earnings

$$\boldsymbol{ForwardPriceToEarnings} = \frac{Price_{Share}}{ForwardEarnings} \quad \text{(Calculation 7)}$$

Calculation Rule 7: Price to Earnings Relative

PE relative$_{\text{Share:Market}}$ and PE relative$_{\text{Share:Market}}$ take into account the fact that market or sectors move in cycles, that is up and down, and underlying shares may move in line with these cycles. PE relative ratios allow an investor to identify changes in the share price that are stock specific as opposed to market or sector related.

$$\boldsymbol{PErelative}_{\boldsymbol{Share:Sector}} = \frac{HistoricPriceToEarnings_{Share}}{HistoricPriceToEarnings_{Sector}} \qquad \text{(Calculation 8)}$$

$$\boldsymbol{PErelative}_{\boldsymbol{Share:Market}} = \frac{HistoricPriceToEarnings_{Share}}{HistoricPriceToEarnings_{Market}} \qquad \text{(Calculation 9)}$$

Calculation Rule 8: Return on Equity

$$\boldsymbol{Return\ on\ Equity} = \frac{Net\ Income}{Total\ Shareholder's\ Equity} \qquad \text{(Calculation 10)}$$

Calculation Rule 9: Cost of Equity

$$\boldsymbol{Equity\ Risk\ Premium} = Market\ Rate\ of\ Return - Risk\ Free\ Rate\ of\ Return \qquad \text{(Calculation 11)}$$

$$\boldsymbol{Cost\ of\ Equity} = Risk\text{-}free\ rate\ of\ return + (Share\ Beta \times Equity\ Risk\ Premium) \qquad \text{(Calculation 12)}$$

Calculation Rule 10: Relative Debt to Equity

$$\boldsymbol{Debt/Equity} = \frac{Total\ Liabilities}{Total\ Shareholder's\ Equity} \qquad \text{(Calculation 13)}$$

$$\boldsymbol{Relative\ Debt\ to\ Equity} = \frac{Debt/Equity_{Share}}{Debt/Equity_{Industry}} \qquad \text{(Calculation 14)}$$

4.5.2. Threshold Evaluation Component (TEC)

This rule set is necessary to evaluate the continuous values of the **Factor** class against thresholds and transform these values into discrete states which are then used by the Bayesian network to make decisions. The values are evaluated against variable thresholds as stipulated by the **Threshold** class and may be altered by a user of the INVEST system.

An instance and its corresponding value for the **MarginOfSafety** class was specified to illustrate its purpose and use in the **TEC**. A user of the system may choose to utilise this default threshold or change the number of instances and values based on their beliefs and preferences. The threshold used throughout the rule set is as follows:

Class	Instance	Value
MarginOfSafety	[%T]	10%
BetaThreshold	[%B]	10%

TEC Rule 1: NegativeEarnings

If ForwardEarnings < 0, then NegativeEarnings = True
Else:
    NegativeEarnings = False

TEC Rule 2: NegativeShareholdersEquity

If ShareholdersEquity < 0, then NegativeShareholdersEquity = True
Else:
    NegativeShareholdersEquity = False

TEC Rule 3: SpecifiedBeta

If value of ShareBeta =< BetaThreshold, then SpecifiedBeta = True
Else:
    SpecifiedBeta = False

TEC Rule 4: Acceptable Stock

If NegativeEarnings = False or NegativeShareholdersEquity = False or SpecifiedBeta = False, then InvestmentDecision = UnacceptableStock
Else:
    InvestmentDecision = AcceptableStock

TEC Rule 4: Current PE relative$_{\mathbf{Share:Market}}$ to Historical PE relative$_{\mathbf{Share:Market}}$

A true bottom up approach for a value investor requires that each individual share is evaluated against the market. It may be the case that a sector is expensive relative to the market but individual shares within the sector can be cheap relative to the market. A true bottom up approach ensures that mispricing is identified at the share level and each share is evaluated against the market.

(1) If the current PE relative$_{Share:Market}$ is less than the rolling historical mean PE relative$_{Share:Market}$ for the share by more than [%T], conclude **cheap.**
(2) If the current PE relative$_{Share:Market}$ is less than the rolling historical mean PE relative$_{Share:Market}$ for the share by more than [%T] or more than the rolling historical mean PE relative for the share by less than [%T], conclude **fairValue**.
(3) If the current PE relative$_{Share:Market}$ is more than the rolling historical mean PE relative$_{Share:Market}$ for the share by more than [%T], conclude **expensive.**

TEC Rule 5: Current PE relative$_{\mathbf{Share:Sector}}$ to Historical PE relative$_{\mathbf{Share:Sector}}$

(1) If the current PE relative$_{Share:Sector}$ is less than the rolling historical mean PE relative$_{Share:Sector}$ for the share by more than [%T], conclude **cheap.**
(2) If the current PE relative$_{Share:Sector}$ is less than the rolling historical mean PE relative$_{Share:Sector}$ for the share by more than [%T] or more than the rolling historical mean PE relative for the share by less than [%T], conclude **fair value**.

(3) If the current PE relative$_{Share:Sector}$ is more than the rolling historical mean PE relative$_{Share:Sector}$ for the share by more than [%T], conclude **expensive.**

TEC Rule 6: ForwardPE Current vs. History

(1) If the **ForwardPE** of the share is less than the rolling historical mean by more than [%T], conclude **cheap.**
(2) If the **ForwardPE** of the share is less than the rolling historical mean by less than [%T] or more than the rolling historical mean by less than [%T], conclude **fairValue.**
(3) If the **ForwardPE**of the share is less than the rolling historical mean by more than [%T], conclude **expensive.**

TEC Rule 7: ROE vs. COE

For profitability performance, return on equity (ROE) will bet in favour of firms with high earnings [23]. Experts pointed out that assessing ROE in isolation may be deceptive. It is imperative that ROE be compared to the cost of equity (COE). If ROE is less than the COE, business value is being eroded. If ROE is more than the COE, value is being created within the business. The wider the spread between the two, the higher the quality of the company.

(1) If ROE is greater than COE for the share by more than [%T], conclude **above.**
(2) If ROE is greater than COE for the share by less than [%T] and or less than COE by less than [%T], conclude **EqualTo.**
(3) If ROE is less than COE for the share by more than [%T], conclude **below.**

TEC Rule 8: CAGR vs. Inflation

For growth performance, empirical studies have evaluated the success of betting in favour of firms with the most positive changes in their profitability factors. Examples include the Levin-Graham strategy [14] which requires 5 years of high past earnings growth while other models [13][23] seek firms whose earnings have grown the most over the past 5 years relative to their year t-5 capital base. The INVEST model has selected the three-year forward compound annual growth rate (CAGR) of earnings per share (EPS) where consensus forecasts are available. In the absence of consensus forecasts, the three-year historical CAGR of EPS is used. This CAGR for EPS, forward-looking or historical, is then compared to inflation. Firms with a CAGR which exceeds inflation should be classified as high quality.

If consensus forecasts are available:
CAGR = ForwardEarnings_CAGR
If no consensus forecasts are available:
CAGR = HistoricalEarnings_CAGR

(1) If CAGR is greater than inflation for the share by more than [%T], conclude **above.**
(2) If CAGR is greater than inflation for the share by less than [%T] and or less than inflation by less than [%T], conclude **EqualTo.**
(3) If CAGR is less than inflation for the share by more than [%T], conclude **below.**

TEC Rule 9: Relative Debt to Equity

For financial risk performance the relative debt to equity ratio (Rel DE) was used in the model. Highly-levered firms are riskier and hence an investor should bet against highly-levered firms [23]. A relative ratio was used given the fact that some firms like banks have structurally higher levels of debt given the nature of their business. Banks receive significant deposits from individuals and institutions which are reflected as debt on the balance sheet. These deposits are used to fund the provision of loans to individuals and institutions which are reflected as assets. The equity component for banks is often small relative to the debt and assets on the balance sheet.

The use of relative factors should be considered on a case-by-case basis to ensure that the share evaluation is not based on misleading factors. In this case, the debt to equity of the share is evaluated relative to the average debt to equity ratio for the industry or sector to which the share belongs.

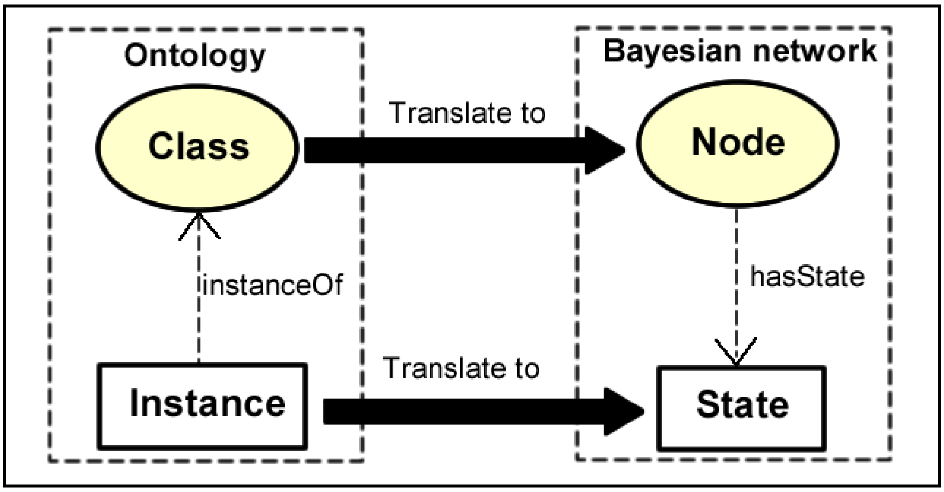
(1) If RelDE is greater than 1 for the share by more than [%T], conclude **above.**
(2) If RelDE is greater than 1 for the share by less than [%T] and or less than 1 by less than [%T], conclude **EqualTo.**
(3) If RelDE is less than 1 for the share by more than [%T], conclude **below.**

4.5.3. Primitive Generator Component for the Bayesian Network (PGC)

The last component contains the transformation algorithm required to generate BN primitives, such as variables, states and probabilities, from the INVEST ontology. The transformation involves the conversion of the **factors** into nodes in the Bayesian network. A node is composed of variables and states which correspond to the classes and their instances respectively. Given that the Threshold Evaluation component has transformed the instances which were continuous values to discrete states for the **Factor** class, the mapping as reflected in **Figure 11** is a direct conversion of the selected factors into nodes and their respective instances into states.

As an example, a concrete sub-class **ForwardPriceToEarnings** of the **Factor** class would be transformed to a node in the **Value Evaluation Bayesian Network (VE_BN)** detailed in Section 4.6.1.2. below. The instances of the **ForwardPriceToEarnings** class, namely **cheap**, **fairValue** and **expensive**, are translated to states of the **ForwardPriceToEarnings**.

Figure 11: Transformation of Invest Ontology primitives into BN primitives



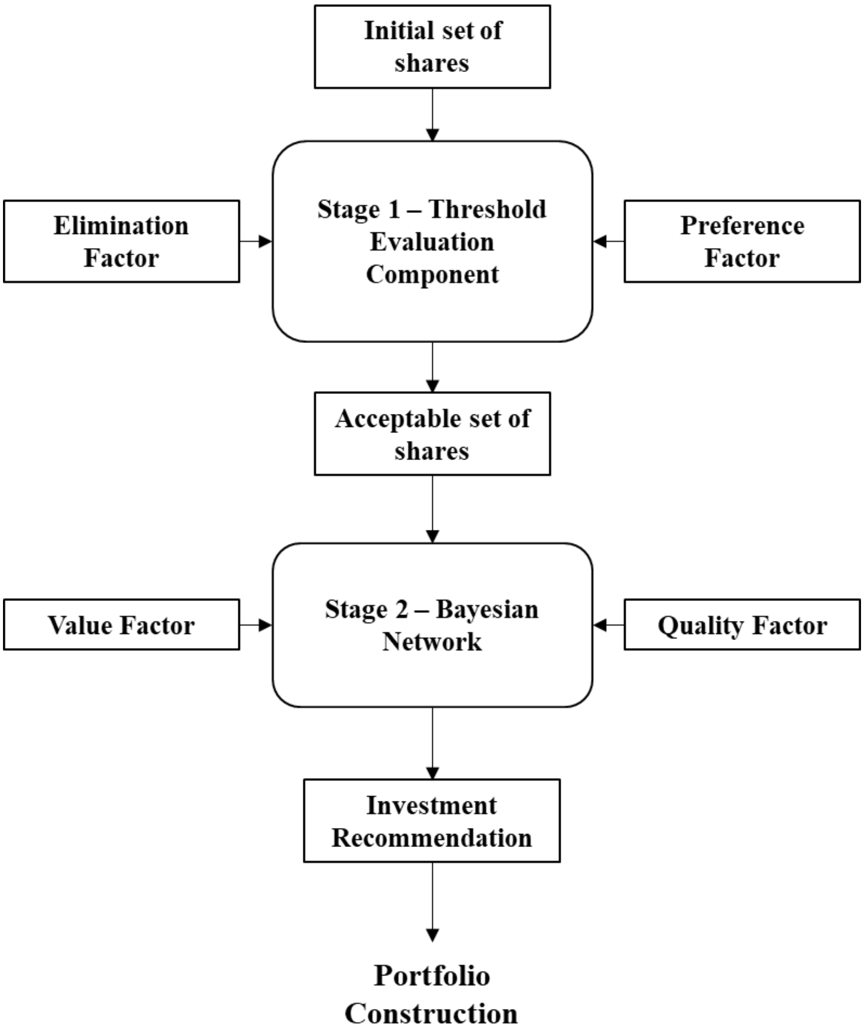
4.6. The Bayesian Network

The Bayesian Network (BN) was designed to support investment professionals in their medium-term share evaluation decisions following a value investing approach in a flexible, practical and realistic manner. Share evaluation comprises two stages as presented in Figure 12 below: (1) the elimination of unacceptable shares and (2) evaluation under uncertainty.

The initial stage eliminates shares that do not meet a minimum set of criteria articulated by investors; this leaves a list of acceptable shares for further evaluation. The initial stage utilizes **EliminationFactors** and **PreferenceFactors** to eliminate shares which do not meet specified investment criteria. The rules for this stage are straightforward and are carried out by the **FC** through the **Threshold Evaluation Component** (see TEC Rule 1 to 4). The result of this procedure is a list of acceptable shares which shall be evaluated through the Bayesian network.

The second stage evaluates the set of acceptable shares utilizing **factors** which the investor believes are influential to or predictive of future share performance.

Figure 12: Process diagram of Share Evaluation mapped to the modelled system



4.6.1. Modelled Bayesian Network for Share Evaluation

A decision network has been employed to represent the decision process; this is an extension of the standard Bayesian network to include utility nodes and decision nodes (see Section 2.8.4. Decision Networks). There are three sequential steps in the decision process which have been modelled using three decision networks as follows:

(1) **Value Evaluation:** Does the share present value relative to price? (Figure 13)
(2) **Quality Evaluation:** Is the share a high, medium or low-quality share? (Figure 14)
(3) **Investment Recommendation:** Is this an investable share or not? (Figure 15)

The first two steps are intermediate decisions that are required as inputs to the final investment recommendation; this is a merged decision which presents a trade-off between value and quality.

4.6.1.1. General Structure

Nature nodes in the decision network (highlighted in "pink") represent the **factors** used to evaluate shares and are used to determine **new beliefs** (in the form of probabilities) as **evidence** is gathered. The decision nodes (highlighted in "blue") correspond to decisions the investor is able to make after evaluating the **evidence**. For example, in reference to Figure 14, the system is able to infer whether the share is high, medium or low-quality based on **evidence** pertaining to three important quality **factors**: CAGRvsInflation; ROEvsCOE and RelDE. These factors are connected to the decision node through **causal links** which reflect what will be known at the time the decision is to be made. The Figure 13 below indicates that the system will know about PE relative$_{Share:Market}$, PE relative$_{Share:Sector}$ and ForwardPE_CurrentvsHistory. The **FutureSharePerformance** node is used in all the decision networks below to reflect the probability of positive, stagnant or negative future share performance which is inferred from the known evidence pertaining to **factors** used to evaluate shares. The probability for this node is determined from historical data which reflects how accurate the **factors** are at predicting future share performance. **Utility nodes** are nodes whose expected value is to be maximized through searching for the **best decision rule** for each of the decision nodes. A **decision rule** indicates which option to choose in making a decision, for each possible condition that may be known when the decision is to be made. A condition is the known evidence (or states) for the **factors** (or variables) in the decision network. The **utility function** for the utility nodes has been established with experts. The combinations of possible states for **FutureSharePerformance** (positive, stagnant, negative) and possible states for **decision nodes** are assigned a utility by experts. For example, if the future share performance is negative but the decision node reflects that the share is cheap the utility for this combination is negative. Netica finds a **decision function** for each decision node by 'solving the net' using the **utility function**.

4.6.1.2. Value Evaluation Bayesian Network (VE_BN)

The **VE_BN** is used to evaluate the value of the share relative to price. The **VE_BN** reflects a two-step decision process with two decision nodes modelled to reflect how investors reason with the respective variables which allow one to evaluate a share's value:

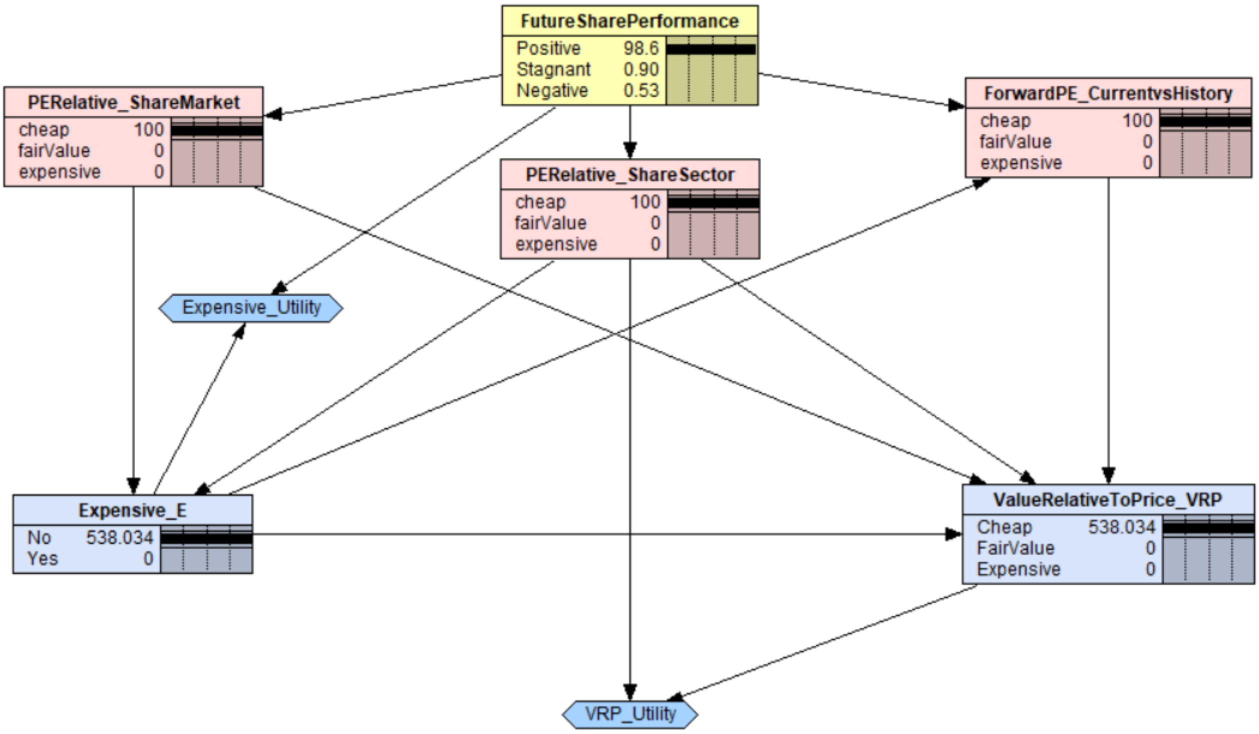
(1) Decision Node 1: Expensive?
Two variables in the Bayesian network; **PE_Relative**$_{\mathbf{Share:Market}}$ and **PE_Relative**$_{\mathbf{Share:Sector}}$; are first evaluated to determine whether the share is expensive or not. If the decision is **"No"**; one continues

to the second decision. If the decision is **"Yes"**, no further evaluation of the share is required with respect to value; the share is expensive.

(2) Decision Node 2: Value Relative to Price?
A third variable, the current **ForwardPE_CurrentvsHistory**, is evaluated to reach a conclusion on the value of the share relative to its price.

Each of the **variables** has a set of states. For example, **ForwardPE_CurrentvsHistory** node has three discrete states, namely "cheap", "fairValue" and "expensive".

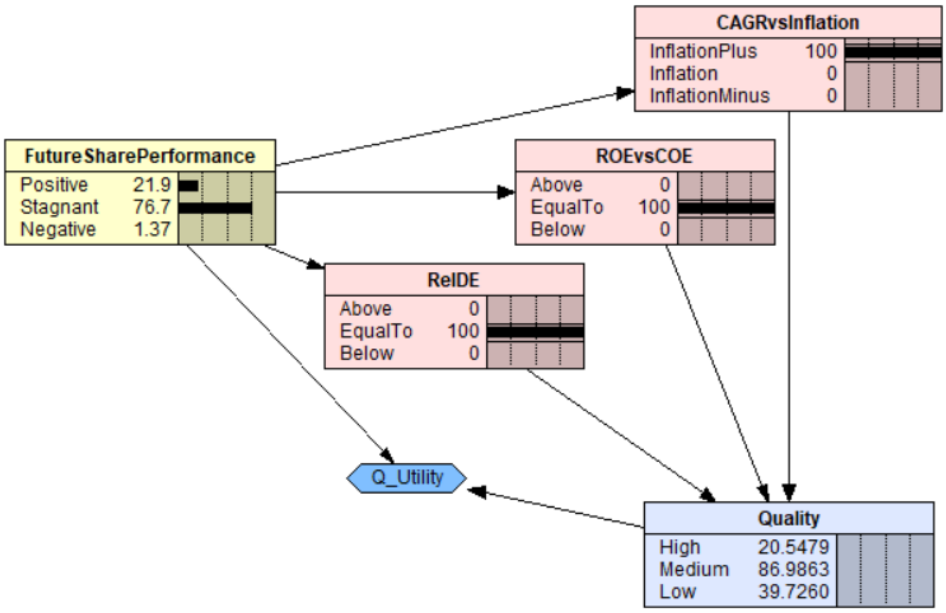
Figure 13: VE_BN Decision Network



4.6.1.3. Quality Evaluation Bayesian Network (QE_BN)

The **QE_BN** is used to evaluate the quality of the share. The **QE_BN** reflects a one-step decision process with one decision node modelled to reflect how investors reason with the respective variables which allow one to evaluate a share's quality. Three variables in the Bayesian network; **Growth_CAGRvsInflation**, **ROEvsCOE** and **Risk_RelDE**; are evaluated using the **decision node** to determine whether a share is high, medium or low quality.

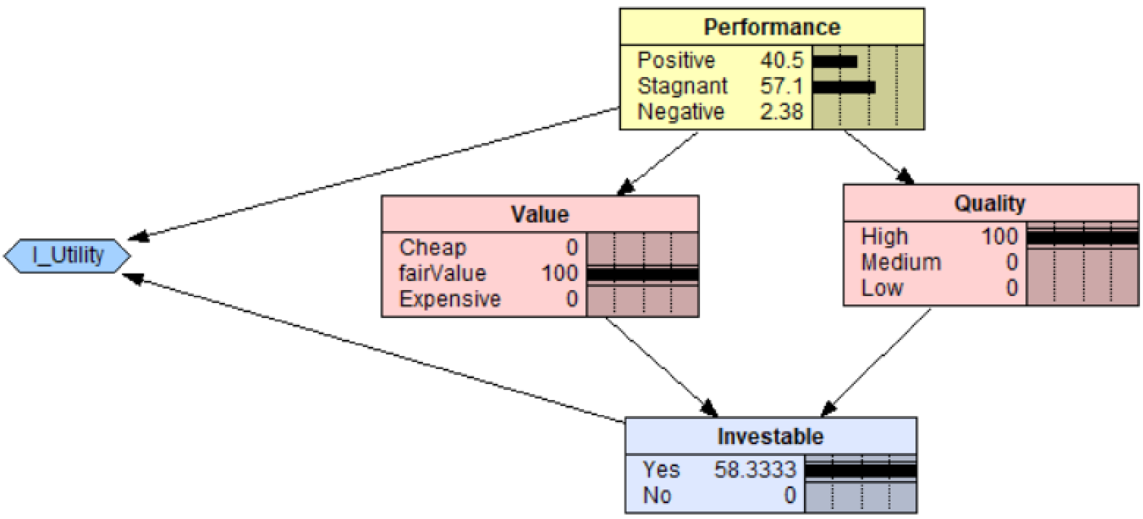
Figure 14: QE_BN Bayesian Network



4.6.1.4. Investment Recommendation Bayesian Network (IR_BN)

The **IR_BN** reflects the decision process an investor must undergo; weighing up whether the price is reasonable and whether the quality of the share high, medium or low. The **IR_BN** is used to provide the investment recommendation for a specific share. The **IR_BN** reflects a one-step decision process with one decision node modelled to reflect how investors reason with the two decision outcomes from **VE_BN** (Value: cheap, fairValue or expensive) and **QE_BN** (Quality: high, medium, low) which are aligned to the variables and states of the variables for **IR_BN**. The **IR_BN** reflects the final trade-off between **Value** and **Quality**. For example, an investor may be willing to pay for a share that is trading at fair value provided it is a high-quality stock but be unprepared to pay for a share trading at fair value should it be a low-quality stock. **Value** and **Quality** are evaluated jointly to reach an investment recommendation.

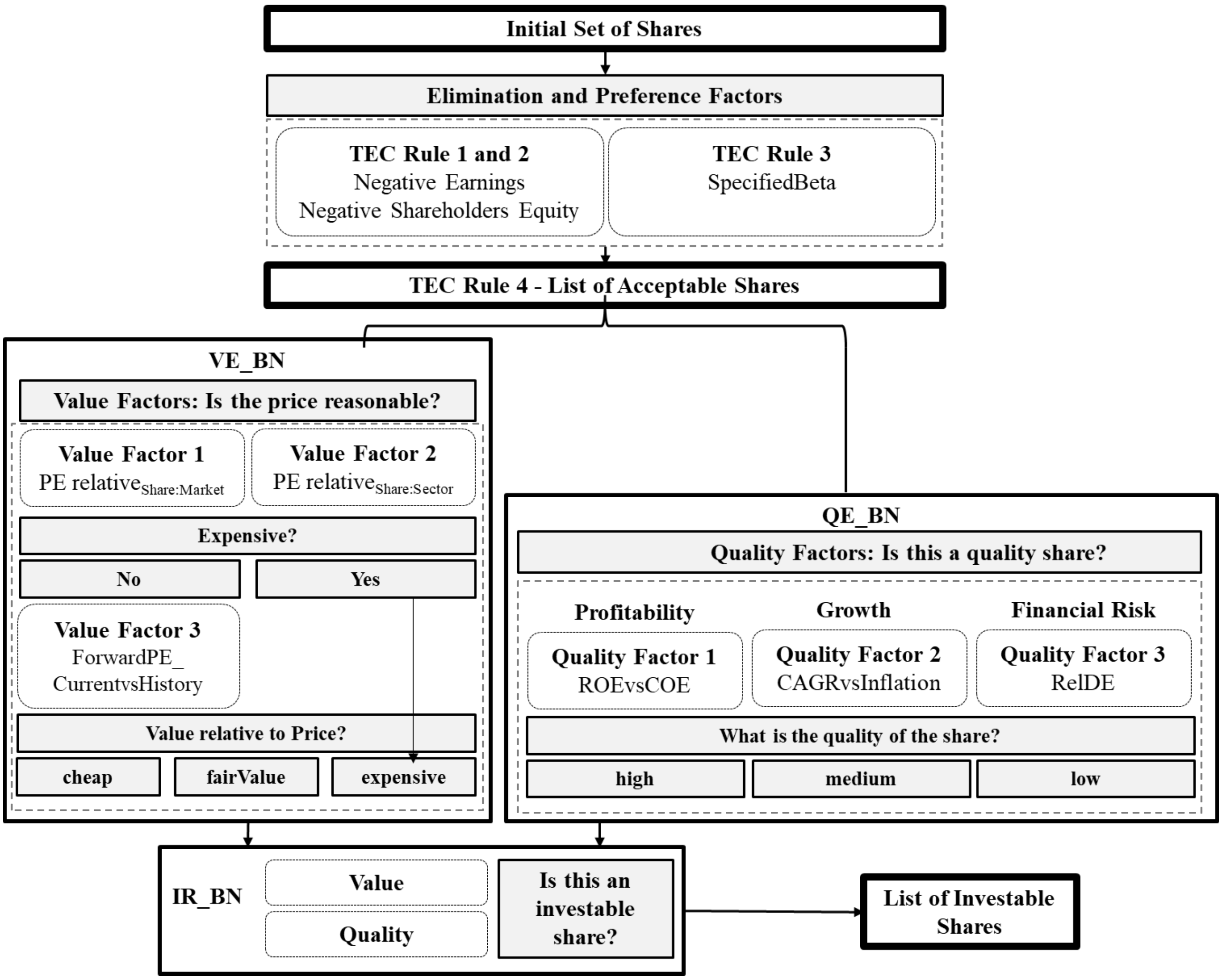
Figure 15: IR_BN Bayesian Network



4.7. Summary

The outcome of the engineering process presented above, is the ontology and Bayesian networks for share evaluation which utilise factors to predict future share performance. The ontology formally represents the key elements required for share evaluation, and extends this for more specific elements pertaining to value investing such that the model becomes useful and concrete. The model is described in a manner that may be shared among human experts and is understandable by machines. Some existing ontologies; FIBO [88] and the SONAR financial ontology [89] were used to supplement this ontology and were extended where necessary. Further to this, a decision model was created for share evaluation under the value investing approach utilising a Bayesian network which draws on the concepts highlighted in the ontology. The share evaluation process is summarised per Figure 16 below.

Figure 16: Summary of the share evaluation process



5. Evaluation of the INVEST System

This chapter describes the evaluation of the ontology and Bayesian network; key components of the INVEST System as presented in Chapter Four. The ontology is evaluated from the aspect of its importance and usefulness within the INVEST System and to the users of this system. The performance of the Bayesian network which captures the decision model for share evaluation following the value investing approach is evaluated in a two-step process: firstly, with experts to refine the decision model and secondly, through back-testing utilizing historical data for shares listed on the JSE.

5.1. Ontology Evaluation

5.1.1. Importance of the INVEST Ontology

While the INVEST ontology (IO) serves the entire INVEST system; the primary purpose of the IO is to support the Bayesian network. The ontology provides support in two primary areas.

Firstly, the ontology facilitates the formulation of the necessary inputs to the Bayesian network (see section 5.1.2. for examples). The ontology utilizes articulates formula which utilizes fundamental data to facilitate the generation of figures and ratios. These figures and ratios, which are continuous values, are evaluated against thresholds, to output qualitative values of factors represented as discrete values which can then serve as inputs to the Bayesian network. The process to transform fundamental data to factors is guided by the clear articulation of all notions for share evaluation and value investing in the ontology. The IO provides the concepts and inputs to the rule set executed by the Financial Calculator. The Bayesian network is constructed utilizing the concepts outlined in the INVEST ontology. The ontology provides clear definitions for the factors employed within the Bayesian network and alleviates any misinterpretations and ambiguity which may exist should it not be in place.

The second area for which the ontology provides support is for the Explanation Facility (EF). The ontology is used to generate explanations to users on how an investment recommendation was reached. The ontology contains the Factor classes and instances of those classes which are fed into the Bayesian network to reach an investment recommendation. The user can query the ontology via the EF to obtain an explanation as to the discrete values for the instances of the Factor classes which are evidence upon which the investment recommendation has been made.

5.1.2. Usage Example

The **VE_BN** Bayesian network described in Chapter 4 is used to illustrate how the ontology supports these two functions.

5.1.2.1. Formulating inputs to the Bayesian Network

As noted above, the ontology is used to formulate the inputs required for the Bayesian network. It should be noted that the terminology of the Bayesian network can be traced back to the ontology. The ontology defines the Factor class which is any ratio, figure or qualitative variable that is believed to be predictive or influential to future share performance. The concrete sub-classes of the Factor class are translated into variables in the Bayesian network. $PE\_Relative_{Share:Market}$, $PE\_Relative_{Share:Sector}$ and

ForwardPriceToEarnings are concrete sub-classes of the Factor class which forms part of the INVEST ontology. These three sub-classes were selected from the INVEST ontology and translated to nature nodes in VE_BN. Each sub-class has discrete instances, cheap; fair value or expensive, which are translated to states in the Bayesian network. The variables and states form the nature nodes which are connected by causal links to decision nodes of the Bayesian network. These three nature nodes are utilised to make the decision: is the share cheap, fair value or expensive relative to its price?

5.1.2.2. Inputs to the Calculator Component

The discrete instances of the nature nodes in the Bayesian networks are created through a series of calculations carried out by the Calculator Component. The IO articulates each input and output for the Calculator Component. Fundamental data (FundamentalData class) is processed by the Calculator Component using defined formula (Formula class). For the PE_Relative$_{\text{Share:Market}}$, PE_Relative$_{\text{Share:Sector}}$ and ForwardPriceToEarnings used in VE_BN there are various concepts required in the IO to formulate these:

- **FundamentalData**: Price; Earnings per share
- **Formula:** HistoricEarnings_GrowthRate; HistoricPricetoEarnings; ForwardEarnings; ForwardPricetoEarnings; PricetoEarningsRelative
- **Classifier:** SectorClassifier
- **InvestmentAsset:** Share; StockMarketIndexFund
- **ModelData:** MarginOfSafety

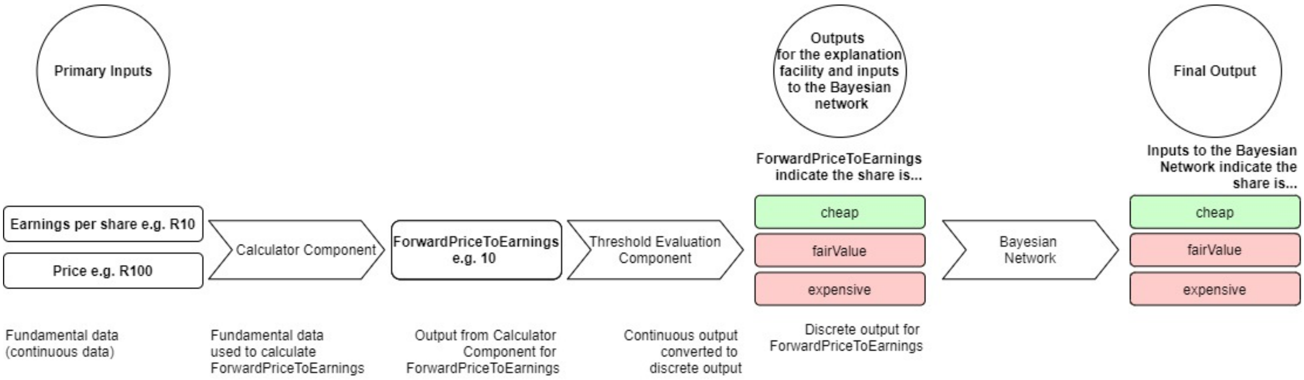
The FundamentalData instances are processed by the Calculator Component to produce the continuous values which are instances of the Formula class. These continuous values are then evaluated against the values for the Threshold class (part of ModelData class) through the Threshold Evaluation Component to create the discrete values which are instances of the Factor class. The function of the Calculator Component is illustrated in Figure 17, which shows the flow of inputs and outputs through the different components in the system.

5.1.2.3. Outputs for the Explanation Facility

The IO is able to facilitate user query and provide explanations for decision outputs from the Bayesian networks. If VE_BN were to conclude that the price of a particular share is “cheap” relative to its value, an investor may want to understand the reason for this decision. The user is able to query the system via the user interface. The EF may present the user with the following output:

PE_Relative$_{\textbf{Share:Market}}$: cheap
PE_Relative$_{\textbf{Share:Sector}}$: fairValue
ForwardPriceToEarnings: cheap

Figure 17: Inputs to the Calculator Component and Outputs to the Explanation Facility



5.2. Expert Evaluation of the Bayesian Network

The Bayesian Network was built based on decision making through the value investing approach as articulated in theory and carried out in practice. The concrete model, inspired by value investing, requires validation. The Bayesian network enables a user to predict future share performance and provides an investment recommendation for a particular share. The evaluation seeks to understand how sound the decision model is and to what extent a value investor could rely on this model to automate decision making.

5.2.1. Experimental Design

The system was designed iteratively with continuous engagement with domain experts. A range of scenarios regarding the value of a share were crafted to iteratively design and evaluate the model. The scenarios contain all possible combinations of factors to ensure that all boundary combinations are included in the evaluation. The different scenarios and expert feedback is shown in Tables 22, 23 and 24.

Four domain experts (DE1 – DE4) from the investment industry and specifically, value investing professionals, participated in the evaluation of the system. The experts had practical experience with value investing and the necessary qualifications to be considered experts (see Table 21 below). However, the experts did not have any experience with Bayesian networks.

Table 21: Qualifications and Years of Experience of Domain Experts

	Qualifications	**Years of Experience**
DE1	CA(SA), CFA	5
DE2	CA(SA), CFA	8
DE3	PGDT Acc Sci, CA (SA), CFA	5
DE4	CA(SA)	7

For each scenario, it was noted whether the results were as expected. Any deviations between the model decision and experts' decisions were investigated and corrected where possible or appropriate. Several rounds of feedback and interactions, between the experts, the modeler and the system, ensure a satisfying convergence is reached between the expert's decisions and the model's decisions. The exercise also allowed the domain expert to re-examine their decisions when it deviated from the models. The feedback from the domain experts were divided into three categories. These categories are outlined below:

- **Feedback A:** the decision maker does not accept the suggestions of the model. A review of the current evaluation factors and/or their weighting within the decision process must ensue.
- **Feedback B:** the decision maker accepts the decision of the model and reviews their manual decision process.
- **Feedback C:** the decision maker accepts the model structure and current evaluation factors but proceeds to modify their stipulated utility based on the combination of decision and outcome to eliminate any further inconsistencies.

5.2.2. Results

The scenario analysis carried out is presented below. This was done for each step of the decision process, made up of two pre-step decisions as per (1) and (2) below, following a final decision as presented in step (3).

(1) **Value Evaluation:** Does the share present value relative to the share price? (Table 22)
(2) **Quality Evaluation:** Is the share high, medium or low quality? (Table 24)
(3) **Investment Recommendation:** Is this share investable or not? (Table 23)

In the tables below, the differences (see cells in pink) are highlighted and were evaluated with experts and where appropriate, the conditional probabilities and utility function of the Bayesian network were adjusted to reduce the differences between the model and experts' decisions (see cells in orange) and ensure that the refined model decisions are implemented.

Table 22 presents the combination of the three **factors** (or **variables**) as modelled in Value Evaluation Bayesian Network (**VE_BN**) which represents the decision model to evaluate a share's value relative to price. There are 24 possible scenarios under the decision hierarchy. The algorithm for the decision hierarchy for **VE_BN** involves a two-step evaluation:

(1) **Step 1:**
   - The model reasons with the states for two variables; namely **PE relative$_{\mathbf{Share:Market}}$** and **PE relative$_{\mathbf{Share:Sector}}$** to evaluate whether the share is expensive or not.
   - If the model concludes that the share is expensive after evaluation of these two factors, then this decision is carried forward.
   - If the model concludes that the share is not expensive after the evaluation of these two factors, then the model proceeds to Step 2.
   - Step 1 serves as a high-level filter for shares that are expensive relative to the market and the sector.

(2) **Step 2:**
   - The model reasons with the states for a third variable; **ForwardPE_CurrentvsHistory** to evaluate whether the share is cheap, fair value or expensive relative to its share price history.

The decision from **VE_BN** was compared to the domain experts' decisions and where appropriate, the model was adjusted such that a refined model was developed from the expert evaluation. Certain decisions were adjusted based on the consensus feedback of experts. For instance, when PE $\text{relative}_{\text{Share:Market}}$ and PE $\text{relative}_{\text{Share:Sector}}$ are both fairValue, the original model stipulated **Yes** for the **Expensive?** decision. This was adjusted based on expert consensus to **No.** In addition, boundary conditions which presented difficulty when creating the decision model were able to be effectively evaluated by experts. For instance, when PE $\text{relative}_{\text{Share:Market}}$ is **Expensive** but PE $\text{relative}_{\text{Share:Sector}}$ is **Cheap** the original model stipulated **Yes** for the **Expensive?** decision. Experts indicated this was punitive and could result in certain shares which are **Cheap** being eliminated from the model too early on.

Table 22: Scenario Analysis for Value Evaluation

			Expensive?							**Value relative to price?**					
Scenario	**ShareMarket _PERelative**	**ShareSector _PERelative**	**Original Model**	**DE1**	**DE2**	**DE3**	**DE4**	**Refined Model**	**Share_PE _CurrentVsHistory**	**Original Model**	**DE1**	**DE2**	**DE3**	**DE4**	**Refined Model**
1a	Cheap	Cheap	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Cheap	**Cheap**	Cheap	Cheap	Cheap	Cheap	**Cheap**
1b	Cheap	Cheap	**Yes**	Yes	Yes	Yes	Yes	**Yes**	fairValue	**Cheap**	Cheap	fairValue	Cheap	Cheap	**Cheap**
1c	Cheap	Cheap	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Expensive	**Expensive**	fairValue	Expensive	fairValue	fairValue	**fairValue**
2a	Cheap	fairValue	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Cheap	**Cheap**	Cheap	Cheap	Cheap	Cheap	**Cheap**
2b	Cheap	fairValue	**Yes**	Yes	Yes	Yes	Yes	**Yes**	fairValue	**fairValue**	fairValue	fairValue	fairValue	fairValue	**fairValue**
2c	Cheap	fairValue	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Expensive	**Expensive**	fairValue	Expensive	fairValue	fairValue	**fairValue**
3a	Cheap	Expensive	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Cheap	**Cheap**	Cheap	Cheap	Cheap	Cheap	**Cheap**
3b	Cheap	Expensive	**Yes**	Yes	Yes	Yes	Yes	**Yes**	fairValue	**fairValue**	fairValue	fairValue	fairValue	fairValue	**fairValue**
3c	Cheap	Expensive	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Expensive	**Expensive**	Expensive	Expensive	Expensive	Expensive	**Expensive**
4a	fairValue	Cheap	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Cheap	**Cheap**	Cheap	Cheap	Cheap	Cheap	**Cheap**
4b	fairValue	Cheap	**Yes**	Yes	Yes	Yes	Yes	**Yes**	fairValue	**Cheap**	Cheap	fairValue	Cheap	Cheap	**Cheap**
4c	fairValue	Cheap	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Expensive	**Expensive**	fairValue	Expensive	fairValue	fairValue	**fairValue**
5a	fairValue	fairValue	**No**	Yes	Yes	Yes	Yes	**Yes**	Cheap	**Cheap**	Cheap	Cheap	Cheap	Cheap	**Cheap**
5b	fairValue	fairValue	**No**	Yes	Yes	Yes	Yes	**Yes**	fairValue	**fairValue**	fairValue	fairValue	fairValue	fairValue	**fairValue**
5c	fairValue	fairValue	**No**	Yes	Yes	Yes	Yes	**Yes**	Expensive	**Expensive**	Expensive	Expensive	Expensive	Expensive	**Expensive**
6	fairValue	Expensive	**Yes**	Yes	No	No	No	**No**	-	**Expensive**	-	-	-	-	**Expensive**
7a	Expensive	Cheap	**No**	Yes	Yes	Yes	Yes	**Yes**	Cheap	**Cheap**	Cheap	Cheap	Cheap	Cheap	**Cheap**
7b	Expensive	Cheap	**No**	Yes	Yes	Yes	Yes	**Yes**	fairValue	**fairValue**	fairValue	fairValue	fairValue	fairValue	**fairValue**
7c	Expensive	Cheap	**No**	Yes	Yes	Yes	Yes	**Yes**	Expensive	**Expensive**	Expensive	Expensive	Expensive	Expensive	**Expensive**
8	Expensive	fairValue	**No**	Yes	No	No	Yes	**No**	-	**Expensive**	-	-	-	-	**Expensive**
9	Expensive	Expensive	**No**	No	No	No	No	**No**	-	**Expensive**	-	-	-	-	**Expensive**

Table 24 below presents the combination of the three **factors** (or **variables**) as modelled in Quality Evaluation Bayesian Network (**QE_BN**) which represents the decision model to evaluate a share's quality. There are 27 possible scenarios under the decision hierarchy. The algorithm for the decision hierarchy for **QE_BN** involves a one-step evaluation:

- The model reasons with the states for three factors (or variables); namely **CAGRvsInflation**, **ROEvsCOE** and **RelDE** to evaluate whether the share is high, medium or low quality.

More differences between the model decision and experts' decisions after reasoning with the quality factors were apparent given that the decision is more nuanced. From the 27 scenarios, differences in the decisions arose on 11 of the scenarios. An example illustrates the reason for the differences: If **ROEvsCOE** is **Below** and **RelDE** is **EqualTo** while **CAGRvsInflation** is **InflationPlus**, these variables must be carefully considered. Historical growth that is inflation plus signals high quality, while a debt to equity ratio in line with the industry reflects medium quality and ROE that is lower than the COE reflects low quality. After reasoning with different experts, it was decided **Medium** quality was appropriate but it is not surprising that **Low** or **High** quality may be selected given the conflicting signals of the factors evaluated.

Table 23 presents the combination of the two **factors** (or **variables**) as modelled in the Investment Recommendation Bayesian Network (**IR_BN**) which represents the decision model to evaluate whether a share is investable or not. There are 9 possible scenarios under the decision hierarchy. The algorithm for the decision hierarchy for **IR_BN** involves a one-step evaluation:

- The model reasons with the states for two factors (or variables); namely **Value** and **Quality** to decide whether the share is investable or not.

Table 23 reflects that there is consensus between the model decision and experts' decisions. **Table 22** and **Table 24** reflect that reasoning with different **factors** to make a decision on **Value** and **Quality** are difficult but once the states of these two variables are known, the merged decision as dealt with through **IR_BN** becomes simple.

Table 23: Scenario Analysis for Investable Evaluation

Scenario	Outcome_Value	Outcome_Quality	Original Model	DE1	DE2	DE3	DE4	Refined Model
1	Cheap	High	Yes	Yes	Yes	Yes	Yes	Yes
2	Expensive	High	No	No	No	No	No	No
3	fairValue	High	Yes	Yes	Yes	Yes	Yes	Yes
4	Cheap	Medium	Yes	Yes	Yes	Yes	Yes	Yes
5	Expensive	Medium	No	No	No	No	No	No
6	fairValue	Medium	No	No	No	No	No	No
7	Cheap	Low	No	No	No	No	No	No
8	Expensive	Low	No	No	No	No	No	No
9	fairValue	Low	No	No	No	No	No	No

Table 24: Scenario Analysis for Quality Evaluation

Scenario	ROEvsCOE	Risk_RelDE	Growth_CAGRvsInflation	What is the quality of the stock?					
				Original Model	**DE1**	**DE2**	**DE3**	**DE4**	**Refined Model**
1	Below	Below	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
2	Above	Above	InflationPlus	**High**	Medium	Medium	Medium	Medium	**Medium**
3	EqualTo	EqualTo	Inflation	**Medium**	Medium	Medium	Medium	Medium	**Medium**
4	Below	EqualTo	Inflation	**Low**	Low	Low	Low	Low	**Low**
5	Below	Above	InflationPlus	**Low**	Low	Medium	Low	Medium	**Low**
6	EqualTo	Above	InflationPlus	**Medium**	Medium	Medium	Medium	Medium	**Medium**
7	EqualTo	Below	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
8	Below	Below	Inflation	**Low**	Low	Low	Low	Low	**Low**
9	Above	Above	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
10	Above	Below	InflationPlus	**High**	High	High	High	High	**High**
11	EqualTo	Above	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
12	EqualTo	Below	Inflation	**Medium**	Medium	Medium	Medium	Medium	**Medium**
13	Below	EqualTo	InflationPlus	**Low**	High	Medium	Medium	High	**Medium**
14	EqualTo	EqualTo	InflationPlus	**Medium**	High	Medium	High	Medium	**Medium**
15	Below	EqualTo	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
16	EqualTo	Above	Inflation	**Medium**	Low	Medium	Low	Low	**Low**
17	Below	Below	InflationPlus	**Low**	Medium	Low	Medium	Medium	**Medium**
18	EqualTo	EqualTo	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
19	Above	EqualTo	Inflation	**Medium**	Medium	High	Medium	High	**Medium**
20	Above	Above	Inflation	**High**	Medium	Medium	Medium	Medium	**Medium**
21	Above	EqualTo	InflationMinus	**Low**	Medium	Medium	Medium	Medium	**Medium**
22	Below	Above	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
23	Above	EqualTo	InflationPlus	**High**	High	High	High	High	**High**
24	Above	Below	InflationMinus	**Low**	Medium	Low	Low	Medium	**Medium**
25	Below	Above	Inflation	**Low**	Low	Low	Low	Low	**Low**
26	Above	Below	Inflation	**Medium**	High	High	Medium	High	**High**
27	EqualTo	Below	InflationPlus	**Medium**	High	Medium	High	High	**High**

5.2.3. Analysis of Evaluation

While a sample of four experts could be considered too small, it has been shown that it is possible to get reliable results from a sample of only 8-12 users [91]. It was felt that the sample of four persons provides a positive indication on the extent to which the Bayesian networks are able to represent and reason with expert knowledge. The evaluation indicated that submitting a draft of the final model to expert evaluation led to further clarity on how to reason with different combinations of factors; the result of which was a more refined model which could be subjected to further testing. In future, more extensive evaluation with more experts should be undertaken.

5.3. Back-testing Evaluation of the Performance of the Bayesian Network

Following the evaluation with experts, the refined model was validated through back-testing the decision models, more specifically the Bayesian networks. This process is detailed below.

5.3.1. Experimental Design

Scope

The Bayesian network has been designed to evaluate shares against their historical relationship to the market and the sector or industry grouping to which they belong. The classification of shares on the JSE at industry and sector level is carried out according to the Global Industry Classification Benchmark (GICS) which is then used to formulate the industry or sector indices. A share index represents a grouping of a set of companies on the stock market that helps investors compare current price levels with past prices to evaluate that grouping's performance. Investors can invest in a share index by buying an index fund, which is generally structured as a mutual fund or exchange traded fund and 'tracks' the index. Two indices, which are tracked by index funds and trade on the JSE, were selected to validate the decision model of the Bayesian networks:

- General Industrials (JGIND) is a sector index
- Consumer Services (JCSEV) is an industry index

The shares for General Industrials and Consumers Services are presented in **Table 26** and **Table 27** respectively. The indices were selected given the sufficient size of the share groupings relative to other indices. Each share within the selected indices was weighted equally (1 share). The selected indices reflect a passive investment strategy and serve as a benchmark to compare the performance of an actively managed portfolio against. The shares within each index represented the investment universe to be evaluated by the three Bayesian networks (**VE_BN**, **QE_BN** and **IR_BN**) to decide whether that particular share was investable or not. Each share that was deemed investable according to **IR_BN** was added to the Investable Portfolio (IP). The performance of the **IP** is to be compared against the respective benchmark.

Data Set

The two selected indices guided the shares to be selected as part of the investment universe. From the original selection of 44 shares, 8 were excluded since they started trading on the JSE after the validation period began (see Table 25). The final data set consisted of 36 shares that were listed on the Johannesburg Stock Exchange (JSE) between 2012 and 2018.

Table 25: Sector Indices selected and number of companies in each

No	Sector Class	Number of Companies	Included in Benchmark Index
1	General Industrials	19	17
2	Consumer Services	25	19
	Total	**44**	**36**

Table 26: Shares within the General Industrials Index

No	Stock Code	Company
1	ADH	Advtech Ltd
2	CLH	City Lodge Hotels Ltd
3	CLS	Clicks Group Ltd
4	COH	Curro Holdings Ltd
5	CSB	Cashbuild Ltd
6	FBR	Famous Brands Ltd
7	ITE	Italtile Ltd
8	LEW	Lewis Group Ltd
9	MRP	Mr Price Group Ltd
10	MSM	Massmart Holdings Ltd
11	PIK	Pick n Pay Stores Ltd
12	SHP	Shoprite Holdings Ltd
13	SPP	SPAR Group Ltd
14	SUI	Sun International Ltd
15	SUR	Spur Corp Ltd
16	TFG	The Foschini Group Ltd
17	TRU	Truworths International Ltd
18	TSG	Tsogo Sun Gaming Ltd
19	WHL	Woolworths Holdings Ltd
20	MCG	MultiChoice Group
21	DCP	Dis-Chem Pharmacies Ltd
22	TGO	Tsogo Sun Hotels Ltd
23	PPH	Pepkor Holdings Ltd
24	MTH	Motus Holdings Ltd
25	SDO	Stadio Holdings Ltd

Table 27: Shares within the Consumer Services Index

No	Stock Code	Company
1	AFT	Afrimat Ltd
2	BAW	Barloworld Ltd
3	BVT	Bidvest Group Ltd
4	GND	Grindrod Ltd
5	HDC	Hudaco Industries Ltd

No	Stock Code	Company
6	IPL	Imperial Logistics Ltd
7	IVT	Invicta Holdings Ltd
8	KAP	KAP Industrial Holdings Ltd
9	MPT	Mpact Ltd
10	MUR	Murray & Roberts Holdings Ltd
11	NPK	Nampak Ltd
12	PPC	PPC Ltd
13	RBX	Raubex Group Ltd
14	RLO	Reunert Ltd
15	SPG	Super Group Ltd
16	TRE	Trencor Ltd
17	WBO	Wilson Bayly Holmes-Ovcon Ltd
18	CTK	Cartrack Holdings Ltd
19	TXT	Textainer Group Holdings Ltd

The following decisions were made with respect to the evaluation of the system:

- Dividends will not be taken into account during the evaluation period to evaluate the model.
- The recommendation from the Bayesian network system does not extend to portfolio construction and therefore does not suggest the weight for each stock within the Investable Portfolio (IP). The IP is thus composed of one share for each stock in the investable set.
- For calculation of relative ratios, namely relative debt-to-equity and PE relative to sector or industry, the shares were grouped into their sector or industry with Customers Services and General Industrials comprising 19 and 17 shares respectively.

Investment Algorithm

A share is evaluated based on its most recent financial data on the 1st of January each year. As an example, the financial statement data for 2012 will be evaluated on the 1st of January 2013. For some metrics, such as historical growth of earnings per share, three years of the preceding historical financial data is required for evaluation. In this example, 2012, 2011 and 2010 financial data would be required. The length of the investment period is one year since the proposed Bayesian network is developed to support medium to long term decisions. The one-year period is from the 1st of January to the 31st of December. On the 1st of January each year, it is assumed that any shares that are investable and not part of the prior year **IP** are bought while any shares that are no longer investable are sold out of the **IP** on this day. The evaluation period for this experiment extends for the company financial year-ends from 2012 to 2017.

Portfolio Performance Evaluation

The performance of the proposed Bayesian network is measured for the validation period by using return and risk-adjusted return measures; discussed in detail below. A key expectation for an active portfolio manager is to derive above-average returns for a given risk class. Investment skill of a portfolio manager is defined as the ability to outperform an appropriate benchmark consistently over time. Thus, a portfolio manager strives for returns in excess of the benchmark; referred to as the active return. That being said, every portfolio manager's active return, regardless of skill, will be positive in some periods and negative in others. A skilful manager should produce a larger active return more frequently than their peers and on average, generate more positive active returns.

Three return metrics detailed in Figure 18 below are used to measure portfolio performance: annual return, average annual return and compound return. Average annual return takes the sum of the historical returns over the investment period and divides this by the number of years. Compound return is the rate of return, expressed as a percentage, that represents the cumulative effect that a series of gains or losses has on an original amount of capital over a period of time. A compound return is expressed in annual terms, meaning that the percentage that is reported represents the annualised rate at which capital has compounded over time.

Figure 18: Return Metrics

$$\text{Annual Return} = \frac{\text{Portfolio Value}_{\text{Year}+1}}{\text{Portfolio Value}_{\text{Year}}} - 1 \qquad \text{(Formula 1)}$$

$$\text{Compound Return} = \left(\frac{\text{Portfolio Value}_{\text{Year}+\text{N}}}{\text{Portfolio Value}_{\text{Year}}}\right)^{1/\text{N}} - 1 \qquad \text{(Formula 2)}$$

$$\text{Average Annual Return} = \frac{\left(\frac{\text{Portfolio Value}_{\text{Year}+\text{N}}}{\text{Portfolio Value}_{\text{Year}}}\right)}{\text{N}} \qquad \text{(Formula 3)}$$

The rate of return is not the only measure of portfolio performance. Minimising risk within the portfolio is crucial since it affects the volatility of returns. Researchers have developed composite portfolio performance measures that measure portfolio returns on a risk-adjusted basis. The *Treynor ratio* and *Sharpe ratio* are examples of risk-adjusted return metrics and have been employed in this study to compare the performance of the portfolios.

Treynor [92] proposed two components of risk; risk produced by general market fluctuations (β) and risk resulting from unique fluctuations (unsystematic risk) in the portfolio of shares [51]. Treynor's ratio is a measurement of the returns earned in excess of that which could have been earned on an investment that has no diversifiable risk (for example, a completely diversified portfolio or a government bond) per unit of systematic risk assumed. Thus, the Treynor ratio relates excess return over the risk-free rate to the additional risk taken; however, systematic risk is used instead of total risk. The higher the Treynor ratio, the better the performance of the portfolio under analysis.

The Sharpe ratio also measures the performance of an investment compared to a risk-free asset, after adjusting for its risk. It is defined as the difference between the returns of the investment and the risk-free return, divided by the standard deviation of the investment which represents the portfolios volatility. The higher the Sharpe ratio, the better the performance of the portfolio under analysis.

Figure 19: Risk-adjusted Return Metrics

$$\text{Treynor Ratio} = \frac{r_p - r_f}{\beta_p}$$

where:

r_p = Portfolio return

r_f = Risk-free rate

β_p = Beta of the portfolio

$$Sharpe\ Ratio = \frac{R_p - R_f}{\sigma_p}$$

where:

R_p = return of portfolio

R_f = risk-free rate

σ_p = standard deviation of the portfolio's excess return

In summary, the **Annual Return, Average Annual Return** and **Compound Annual Return** will be used as pure return metrics for performance evaluation while the **Treynor Ratio** and **Sharpe Ratio** will be used to incorporate risk into the performance evaluation.

5.3.2. Results

The **IR_BN** recommended the following companies for inclusion in the **IP for General Industrial shares (IP.JGIND)** as presented in **Table 27** and for inclusion in the **IP for Customer Service shares (IP.JCSEV)** as presented in **Table 28** for the next 12-month period from the 1st of January to the 31st of December.

Table 28: IP.JGIND from 2012 to 2018

		Investable Shares (Yes/No)					
No	**Stock Code**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
1	AFT	**Yes**	No	No	No	No	No
2	BAW	**Yes**	**Yes**	**Yes**	No	**Yes**	No
3	BVT	**Yes**	No	No	No	**Yes**	**Yes**
4	GND	No	No	No	No	No	No
5	HDC	No	No	No	**Yes**	**Yes**	**Yes**
6	IPL	No	No	No	No	No	No
7	IVT	No	No	No	No	No	No
8	KAP	No	**Yes**	**Yes**	No	No	No
9	MPT	**Yes**	**Yes**	**Yes**	**Yes**	No	No
10	MUR	No	No	**Yes**	**Yes**	**Yes**	No
11	NPK	**Yes**	No	No	No	No	No
12	PPC	No	No	No	No	No	No
13	RBX	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No
14	RLO	**Yes**	No	**Yes**	**Yes**	**Yes**	**Yes**

		Investable Shares (Yes/No)					
No	**Stock Code**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
15	SPG	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No
16	TRE	**Yes**	**Yes**	No	No	No	No
17	WBO	No	No	**Yes**	**Yes**	**Yes**	No
18	CTK	Excluded since they started trading on the JSE after the validation period began.					
19	TXT						
Total Investable Shares		**9**	**6**	**8**	**7**	**8**	**3**

Table 29: IP.JCSEV from 2012 to 2018

		Investable Shares (Yes/No)					
No	**Stock Code**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
1	ADH	No	No	No	No	No	No
2	CLH	No	No	**Yes**	**Yes**	**Yes**	**Yes**
3	CLS	No	No	No	**Yes**	No	No
4	COH	**Yes**	No	**Yes**	**Yes**	**Yes**	No
5	CSB	No	No	No	**Yes**	No	No
6	FBR	No	No	No	No	No	No
7	ITE	**Yes**	**Yes**	No	**Yes**	No	No
8	LEW	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No
9	MRP	No	No	No	No	No	No
10	MSM	No	No	No	**Yes**	No	No
11	PIK	No	No	No	**Yes**	No	No
12	SHP	No	No	No	No	**Yes**	No
13	SPP	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
14	SUI	**Yes**	**Yes**	No	No	No	No
15	SUR	No	No	No	No	No	No
16	TFG	No	No	**Yes**	No	No	No
17	TRU	No	No	No	No	No	**Yes**
18	TSG	**Yes**	No	**Yes**	**Yes**	**Yes**	**Yes**
19	WHL	No	No	No	No	No	**Yes**
20	MCG	Excluded since they started trading on the JSE after the validation period began.					
21	DCP						
22	TGO						
23	PPH						
24	MTH						
25	SDO						
Total Investable Shares		**5**	**4**	**6**	**10**	**6**	**5**

Table 30: Performance Measures

Period	IP.JGIND AR*	Benchmark JGIND AR*	Active Return	IP.JCSEV AR*	BenchmarkJCSEV AR*	Active Return
2012 to 2013	**30.10%**	26.20%	3.90%	**23.30%**	18.04%	5.26%
2013 to 2014	**16.92%**	8.30%	8.62%	**12.73%**	9.28%	3.45%
2014 to 2015	**- 0.38%**	- 9.32%	8.94%	**45.42%**	25.09%	20.33%
2015 to 2016	- 1.96%	**0.04%**	-2.00%	**4.86%**	2.34%	2.52%
2016 to 2017	**20.92%**	14.55%	6.38%	- **0.26%**	- 0.33%	0.06%
2017 to 2018	**13.89%**	- 1.17%	15.06%	**-** 2.01%	**1.19%**	-3.20%
Measure	**IP.JGIND**	**Benchmark JGIND**	**Delta**	**IP.JCSEV**	**BenchmarkJCSEV**	**Delta**
Compound Return	**13.25%**	6.43%	6.82%	**12.90%**	8.87%	4.03%
Average Annual Return	**6.25%**	-0.57%	6.82%	**14.01%**	9.27%	4.74%
Standard Deviation	12.44%	12.70%	-0.26%	17.99%	10.32%	7.67%
Treynor Ratio	**0.10**	-0.01	0.11	**0.13**	0.03	0.10
Sharpe Ratio	**0.50**	-0.04	0.54	**0.39**	0.22	0.17

* Annual Return

The performance of the decision model encapsulated by the Bayesian networks is measured for the validation period by using the return and risk adjusted return measures detailed above. Performances of the portfolios are compared with the benchmark indices selected.

Annual Return

The portfolio annual return depicted in **Table 30** was determined by comparing the acquisition price to the sell price for each stock in the portfolio. The IP.JGIND portfolio outperformed the benchmark index in terms of annual return for the six one-year periods from 2012 to 2018. The IP. JSCEV outperformed the benchmark index in terms of annual return for four of the six one-year periods from 2012 to 2018.

Compound Return

Table 30 demonstrates the performance evaluation results of the proposed expert system in terms of compound return. Both the IP.JGIND and IP.JCSEV yield higher compound returns and average annual returns than the benchmark index yields.

Treynor Ratio

Table 30 also demonstrates the performance evaluation results of the proposed expert system in terms of the Treynor ratio. The Treynor ratio for IP.JGIND was 0.10 and exceeds the BI.JGIND which was -0.01. The Treynor ratio for IP.JCSEV was 0.13 and exceeds the BI.JCSEV. It is important to reiterate that a higher ratio indicates a more favourable risk/return scenario

Sharpe Ratio

Table 30 demonstrates the performance evaluation results of the proposed ES in terms of the Sharpe ratio. The Sharpe ratio for IP.JGIND was 0.50 and exceeds the BI.JGIND which was -0.04. The Sharpe ratio for IP.JCSEV was 0.39 and exceeds the BI.JCSEV which was 0.22. It must be noted that on an absolute basis a Sharpe ratio below 1 is sub-optimal, while a ratio above 1 is acceptable. In this case, both the IP and BI portfolios were below 1 and are therefore sub-optimal on an absolute basis.

5.3.3. Analysis of Evaluation

The evaluation utilised shares in two indices, which represent subsets of shares listed on the JSE, as the initial set of shares to be evaluated. While the results are promising a larger investment universe and sample size may be required to conclude that the results are statistically significant. The General Industrials index and Consumer Services index served as proxies for larger share groupings like the All Share Index (ALSI) on the JSE. The choice to evaluate the Bayesian networks' performance against sector or industry indices is reflective of the manner in which investors analyse shares. Investors engage in contextual fundamental analysis which involves evaluating shares within their sector or industry groupings. That being said, a larger and more diverse benchmark index such as the ALSI may provide more decisive results. The performance of the proposed decision model as presented using three Bayesian networks is superior to the selected benchmark indices over the period from 2012 to 2018 based on return and risk-adjusted return measures. The fact that some well-performing shares within the dataset were not selected in certain years suggests that the model could be improved with the addition of further factors or model refinement to ensure these shares are not missed. The risk-adjusted return measures like Treynor and Sharpe support this view given that on an absolute basis the performance of the model is below the desired levels. This suggest one of two things: (1) the model performance could be refined or (2) the industry or sector of the indices evaluated have inherently poor risk-adjusted return characteristics. A conclusive answer to this could be reached through further testing. In future, the evaluation could be extended to analyse performance of the Bayesian networks over different holding periods. For the purposes of this evaluation, a 12-month holding period was selected in line with the full-year financial results cycle of companies.

5.4. Summary

The evaluation above centres on the key components to the INVEST System: the ontology and Bayesian network. The ontology was briefly evaluated through consideration of its importance and usefulness within the INVEST System. The evaluation of the Bayesian network was more rigorous to ensure the performance of the captured the decision model for share evaluation under the value investing approach was sound. While the evaluation yielded promising results there are two points to note. Firstly, the results from the evaluation should be used to inform future system designs for share evaluation. Secondly, further evaluation should be carried out on the INVEST System to conclude on its performance.

6. Discussion and Conclusion

This research investigates the efficacy of ontologies and Bayesian networks for automating share evaluation focused on prediction of future share performance under the value investing approach. It was found that ontologies and Bayesian networks are promising technologies that are complementary in the construction of predictive models for future share performance. The ontology provided the means to clearly articulate concepts necessary for share evaluation and facilitated the provision of inputs to the Bayesian network. The Bayesian network was able to deal with the uncertainty inherent in share evaluation decision making.

The research highlighted that constructing predictive models for future share performance is challenging. Current knowledge on share evaluation and future return prediction models is embedded in a diverse and growing volume of studies which are frequently ambiguous, confusing and fail to be useful for concrete application in the real world. In addition, many existing studies and current state of the art fail to explicitly consider expert knowledge or investment theory in the creation of their decision models. This research explores a new categorization for factors which are useful in predicting future share performance. This provided a useful framework to construct the future share performance prediction model. Linking factors with evaluation objectives and the underlying investment approach is useful as it allows an investor to select the correct factors for the decision model based on their beliefs and to make explicit the evaluation objectives of the decision model.

Evaluation was carried out to ensure that the expert knowledge is correctly represented and the system design is useful. The initial refinement and evaluation with experts revealed that automated share evaluation provides an investor with an explicit decision framework and reflection point off which to make better decisions. It also highlighted that while experts agreed on the inputs, they disagreed with the model outputs for the boundary conditions where the model is tested at the extreme values. These disagreements do not necessarily negate the usefulness of the model but they represent the view of different market participants and reflect the uncertainty inherent in each investment decision. Future studies may aim to tackle these boundary conditions through the refinement of evaluation factors included in the model or expansion of the state’s pertaining to these evaluation factors. The back-testing results for the Bayesian network are promising; the performance of the proposed decision model was superior to benchmark indices for the evaluation period and investment universe selected. Though the results are positive, several observations suggest the model could be refined further.

Future work on developing share evaluation models can draw on this work as a framework for share evaluation and explore the addition of further factors which may be meaningful for prediction of future share performance. Even though the model focused on the value investing approach, the model can be easily adapted and extended for alternative investment approaches for share evaluation through the addition of more or alternative factors.

7. References

1. Benjamin, G. and Dodd, D.L., 1934. Security analysis. Me Graw Hill Ine, New York.
2. Malkiel, B.G., 2003. The efficient market hypothesis and its critics. Journal of economic perspectives, 17(1), pp.59-82.
3. Malkiel, B.G. and Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. The journal of Finance, 25(2), pp.383-417.
4. Basu, S., 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. The journal of Finance, 32(3), pp.663-682.
5. S. L. Jones and J. M. Netter, n.d. Efficient Capital Markets. The Library of Economics and Liberty.
6. Grossman, S.J. and Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. The American economic review, 70(3), pp.393-408.
7. Kruger, R. and Toerien, F., 2014. The consistency of equity style anomalies on the JSE during a period of market crisis. African Finance Journal, 16(1), pp.1-18.
8. Muller, C. and Ward, M., 2013. Style-based effects on the Johannesburg Stock Exchange: A graphical time-series approach. Investment Analysts Journal, 42(77), pp.1-16.
9. Marks, H., 2018. Mastering the market cycle: getting the odds on your side. Houghton Mifflin Harcourt.
10. Keynes, J.M., 1936. The general theory of interest, employment and money.
11. Williams, J.B., 1938. The theory of investment value (No. HG4521 W48).
12. Lee, C.M., 2014. Value investing: Bridging theory and practice. *China Accounting and Finance Review*, 16(2), pp.1-29.
13. Graham, B., Dodd, D.L.F., Murray, R.F. and Block, F.E., 1988. Graham and Dodd's security analysis. McGraw-Hill Professional.
14. Levin, B., n.d. Levin-Graham Stock Screen presented in Value investing: Bridging theory and practice. *China Accounting and Finance Review*, *16*(2), pp.1-29.
15. Greenblatt, J., 2010. The little book that still beats the market (Vol. 29). John Wiley & Sons.
16. Yunusoglu, M.G. and Selim, H., 2013. A fuzzy rule based expert system for stock evaluation and portfolio construction: An application to Istanbul Stock Exchange. Expert Systems with Applications, 40(3), pp.908-920.
17. Penman, S., 2016. Valuation: The state of the art. Schmalenbach Business Review, 17(1), pp.3-23.
18. Dedi, L. and Giraudon, P., 2013. Valuation and investment profession. UTMS Journal of Economics, 4(2), pp.93-100.
19. Bancel, F. and Mittoo, U., 2012. Valuation practices and CAPM implementation. In 19th Annual Conference of the Multinational Finance Society, Krakow, Poland. Retrieved from http://www. cfasociety. org/switzerland/Lists/Events% 20Calendar/Attachments/31 (Vol. 8).
20. Pinto, J.E., Robinson, T.R. and Stowe, J.D., 2019. Equity valuation: A survey of professional practice. Review of Financial Economics, 37(2), pp.219-233.
21. Beneda, N.L., 2003. Estimating free cash flows and valuing a growth company. Journal of Asset Management, 4(4), pp.247-257.
22. Novy-Marx, R., 2013. The quality dimension of value investing. Rnm. simon. rochester. edu, pp.1-54.
23. Asness, C.S., Frazzini, A. and Pedersen, L.H., 2014. Low-risk investing without industry bets. Financial Analysts Journal, 70(4), pp.24-41.
24. Sloan, R.G., 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings?. Accounting review, pp.289-315.

25. Richardson, S.A., Sloan, R.G., Soliman, M.T. and Tuna, I., 2005. Accrual reliability, earnings persistence and stock prices. Journal of accounting and economics, 39(3), pp.437-485.
26. Beneish, M.D., 1999. The detection of earnings manipulation. Financial Analysts Journal, 55(5), pp.24-36.
27. Beneish, M.D., Lee, C.M. and Nichols, D.C., 2013. Earnings manipulation and expected returns. Financial Analysts Journal, 69(2), pp.57-82.
28. Aouni, B., Doumpos, M., Pérez-Gladish, B. and Steuer, R.E., 2018. On the increasing importance of multiple criteria decision aid methods for portfolio selection. Journal of the Operational Research Society, 69(10), pp.1525-1542.
29. Mimović, P., Stanković, J. and Janković Milić, V., 2015. Decision-making under uncertainty–the integrated approach of the AHP and Bayesian analysis. Economic research-Ekonomska istraživanja, 28(1), pp.868-877.
30. Samaras, G.D., Matsatsinis, N.F. and Zopounidis, C., 2008. A multicriteria DSS for stock evaluation using fundamental analysis. European Journal of Operational Research, 187(3), pp.1380-1401.
31. De Nicola, A., Missikoff, M. and Navigli, R., 2005, August. A proposal for a unified process for ontology building: UPON. In International Conference on Database and Expert Systems Applications (pp. 655-664). Springer, Berlin, Heidelberg.
32. Westrick, L., Du, J. and Wolffe, G., 2015. Building a Better Stockbroker: Managing Big (Financial) Data by Constructing an Ontology-Based Framework. Journal of Information Systems Applied Research, 8(1), p.31.
33. Feilmayr, C. and Wöß, W., 2016. An analysis of ontologies and their success factors for application to business. Data & Knowledge Engineering, 101, pp.1-23.
34. Yang, Y. and Calmet, J., 2005, November. Ontobayes: An ontology-driven uncertainty model. In *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)* (Vol. 1, pp. 457-463). IEEE.
35. Bennett, M., 2011, May. Semantics standardization for financial industry integration. In *2011 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 439-445). IEEE.
36. Guarino, N., Oberle, D. and Staab, S., 2009. What is an ontology? Handbook on Ontologies. *Handbook on Ontologies SE-International Handbooks on Information Systems*, pp.1-17.
37. Guarino, N. and Giaretta, P., 1995. Ontologies and knowledge bases. *Towards very large knowledge bases*, pp.1-2.
38. Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing?. *International journal of human-computer studies*, *43*(5-6), pp.907-928.
39. Borst, W., 1997. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse, Centre for Telematica and Information Technology, University of Twente*. Enschede, technical report.
40. Neches, R., Fikes, R.E., Finin, T., Gruber, T., Patil, R., Senator, T. and Swartout, W.R., 1991. Enabling technology for knowledge sharing. *AI magazine*, *12*(3), pp.36-36.
41. Hendler, J., 2001. Agents and the semantic web. IEEE Intelligent systems, 16(2), pp.30-37.
42. Uschold, M. and Gruninger, M., 1996. Ontologies: Principles, methods and applications. *The knowledge engineering review*, *11*(2), pp.93-136.
43. Kanellopoulos, D., Kotsiantis, S. and Tampakas, V., 2007, September. Towards an ontology-based system for intelligent prediction of firms with fraudulent financial statements. In *2007 IEEE Conference on Emerging Technologies and Factory Automation (EFTA 2007)* (pp. 1300-1307). IEEE.
44. Roussey, C., Pinet, F., Kang, M.A. and Corcho, O., 2011. An introduction to ontologies and ontology engineering. In *Ontologies in Urban development projects* (pp. 9-38). Springer, London.

45. Gomez-Perez, A., Fernández-López, M. and Corcho, O., 2006. *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer Science & Business Media.
46. Fonseca, F., Egenhofer, M., Davis, C. and Câmara, G., 2002. Semantic granularity in ontology-driven geographic information systems. *Annals of mathematics and artificial intelligence*, *36*(1-2), pp.121-151.
47. Fensel, D., van Harmelen, F. and Horrocks, I., 1999. OIL: A standard proposal for the Semantic Web. *On-To-Knowledge deliverable D-0, Vrije Universiteit Amsterdam*.
48. Fernández-López, M., Gómez-Pérez, A. and Juristo, N., 1997. Methontology: from ontological art towards ontological engineering.
49. Uschold, M., 1996, September. Building ontologies: Towards a uni ed methodology. In *16th Annual Conf. of the British Computer Society Specialist Group on Expert Systems, Cambridge, UK*.
50. Fernández-López, M. and Gómez-Pérez, A., 2002. Overview and analysis of methodologies for building ontologies. *The knowledge engineering review*, *17*(2), pp.129-156.
51. Browne, O., Krdzavac, N., O'Reilly, P. and Hutchinson, M., 2017. Semantic Ontologies and Financial Reporting: An Application of the FIBO. In *JOWO*.
52. De Nicola, A., Missikoff, M. and Navigli, R., 2005, August. A proposal for a unified process for ontology building: UPON. In *International Conference on Database and Expert Systems Applications* (pp. 655-664). Springer, Berlin, Heidelberg.
53. Iqbal, R., Murad, M.A.A., Mustapha, A. and Sharef, N.M., 2013. An analysis of ontology engineering methodologies: A literature review. *Research journal of applied sciences, engineering and technology*, *6*(16), pp.2993-3000.
54. Islam, N., Abbasi, A.Z. and Shaikh, Z.A., 2010, June. Semantic Web: Choosing the right methodologies, tools and standards. In *2010 International Conference on Information and Emerging Technologies* (pp. 1-5). IEEE.
55. Horrocks, I., Patel-Schneider, P.F. and Van Harmelen, F., 2003. From SHIQ and RDF to OWL: The making of a web ontology language. *Journal of web semantics*, *1*(1), pp.7-26.
56. Grau, B.C., Kharlamov, E., Kostylev, E.V. and Zheleznyakov, D., 2013, October. Controlled query evaluation over OWL 2 RL ontologies. In *International Semantic Web Conference* (pp. 49-65). Springer, Berlin, Heidelberg.
57. Brickley, D., Guha, R.V. and McBride, B., 2014. Rdf schema 1.1. W3C recommendation. *World Wide Web Consortium, February*.
58. Gaševic, D., Djuric, D. and Devedžic, V., 2009. Ontologies. In *Model Driven Engineering and Ontology Development* (pp. 45-80). Springer, Berlin, Heidelberg.
59. Banerjee, J., 1987. H.-t. Chou, JF Garza, and W. Kim,"Data Model Issues for Object-Oriented Applications,". *ACM Transactions on Office Information Systems*, *5*, pp.3-26.
60. Du, J. and Zhou, L., 2012. Improving financial data quality using ontologies. *Decision Support Systems*, *54*(1), pp.76-86.
61. Kalcheva, K., 2016. *Linked Data adoption and application within financial business processes*. Anchor Academic Publishing.
62. Zhu, H. and Wu, H., 2014. Assessing the quality of large-scale data standards: A case of XBRL GAAP Taxonomy. *Decision Support Systems*, *59*, pp.351-360.
63. Chowdhuri, R., Yoon, V.Y., Redmond, R.T. and Etudo, U.O., 2014. Ontology based integration of XBRL filings for financial decision making. *Decision Support Systems*, *68*, pp.64-76.
64. Li, X., Bao, J. and Hendler, J.A., 2011, April. Fundamental analysis powered by semantic web. In 2011 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr) (pp. 1-8). IEEE.

65. Pease, A., Niles, I. and Li, J., 2002, July. The suggested upper merged ontology: A large ontology for the semantic web and its applications. In *Working notes of the AAAI-2002 workshop on ontologies and the semantic web* (Vol. 28, pp. 7-10).
66. Geerts, G.L. and McCarthy, W.E., 2002. An ontological analysis of the economic primitives of the extended-REA enterprise information architecture. International Journal of Accounting Information Systems, 3(1), pp.1-16.
67. Madar, L., 2014. Ontological Engineering: Creating a Retail Banking Ontology (Created for the Department of Engineering, University of Buffalo).
68. Hoffman, 2014. FIBO is a Sign of Things to Come. *Digital Financial Reporting: Intelligent XBRL-based structured digital financial reporting*.
69. Zaino, J., 2012. Financial Services Industry Sees Operational Value in FIBO. Dataversity.
70. EDM Council, 2013. Financial Industry Business Ontology (FIBO) On Track to Become a Global Financial Technology Standard.
71. Salah, M. and Mohamed, T., 2011. Developing Ontology for Financial Investments "Algeria Case Study". International Journal of Computer Applications, 975, p.8887.
72. Shan, W.S., 2009. *Knowledge driven data mining for causal relationships between news and financial instruments* (Doctoral dissertation, Ph. D. Thesis, City University of Hong Kong).
73. Mitchell, H.B., 2012. *Data fusion: concepts and ideas*. Springer Science & Business Media.
74. Zhang, D. and Zhou, L., 2004. Discovering golden nuggets: data mining in financial application. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *34*(4), pp.513-522.
75. Hu, D., Yan, J., Zhao, J.L. and Hua, Z., 2014. Ontology-based scenario modeling and analysis for bank stress testing. *Decision Support Systems*, *63*, pp.81-94.
76. Rodríguez-González, A., Colomo-Palacios, R., Guldris-Iglesias, F., Gómez-Berbís, J.M. and García-Crespo, A., 2012. FAST: Fundamental analysis support for financial statements. Using semantics for trading recommendations. *Information Systems Frontiers*, *14*(5), pp.999-1017.
77. Demirer, R., Mau, R.R. and Shenoy, C., 2006. Bayesian networks: a decision tool to improve portfolio risk analysis. *Journal of applied finance*, *16*(2), p.106.
78. Russel, S. and Norvig, P., 2013. *Artificial intelligence: a modern approach*. Pearson Education Limited.
79. Tseng, C.C., 2006. Influence diagram for investment portfolio selection. In *Computational Economics: A Perspective from Computational Intelligence* (pp. 62-78). IGI Global.
80. Korb, K.B. and Nicholson, A.E., 2010. *Bayesian artificial intelligence*. CRC press.
81. Cassim, Z., 2016. *Predicting financial distress of JSE-listed companies using Bayesian networks* (Doctoral dissertation, University of Cape Town).
82. Tversky, A. and Kahneman, D., 1989. Rational choice and the framing of decisions. In *Multiple criteria decision making and risk analysis using microcomputers* (pp. 81-126). Springer, Berlin, Heidelberg.
83. Kahneman, D. and Tversky, A., 1982. Variants of uncertainty. *Cognition*, *11*(2), pp.143-157.
84. Tseng, C., Gmytrasiewicz, P.J. and Ching, C., 2001, April. Refining influence diagram for stock portfolio selection. In *Proceedings of the Seventh International Conference of the Society for Computational Economics* (pp. 241-256).
85. Ogundele, O.A., Moodley, D., Pillay, A.W. and Seebregts, C.J., 2016. An ontology for factors affecting tuberculosis treatment adherence behavior in sub-Saharan Africa. *Patient preference and adherence*, *10*, p.669.
86. Riley, G. and Giarratano, J.C., 2005. *Expert systems: principles and programming*. Thomson Course Technology.

87. Badiru, A.B. and Cheung, J., 2002. *Fuzzy engineering expert systems with neural network applications* (Vol. 11). John Wiley & Sons.
88. Council, E., 2016. FIBO. The Financial Industry Business Ontology.
89. Gómez, J.M., García-Sánchez, F., Valencia-García, R., Toma, I. and Moreno, C.G., 2009, June. Sonar: A semantically empowered financial search engine. In *International work-conference on the interplay between natural and artificial computation* (pp. 405-414). Springer, Berlin, Heidelberg.
90. Barra, M.S.C.I., 2009. *Global industry classification standard (gics)*. Technical report, Standard & Poors.
91. Tullis, T.S. and Stetson, J.N., 2004, June. A comparison of questionnaires for assessing website usability. In *Usability professional association conference* (Vol. 1).
92. Treynor, J.L., 2011. *Treynor on institutional investing* (Vol. 402). John Wiley & Sons.
93. Studer, R., Benjamins, V.R. and Fensel, D., 1998. Knowledge engineering: principles and methods. Data & knowledge engineering, 25(1-2), pp.161-197.
94. OWL Working Group, 2009. OWL 2 Web Ontology Language Document Overview: W3C Recommendation 27 October 2009.
95. Fensel, D., 2001. Ontologies. In Ontologies (pp. 11-18). Springer, Berlin, Heidelberg.
96. Kämpgen, B., Weller, T., O'Riain, S., Weber, C. and Harth, A., 2014, May. Accepting the xbrl challenge with linked data for financial data integration. In European Semantic Web Conference (pp. 595-610). Springer, Cham.
97. Damodaran, A., 2013. Valuing financial services firms. Journal of Financial Perspectives, 1(1).

8. Appendix

Appendix 1: OWL representation of the INVEST Ontology


```
<?xml version="1.0"?>
<Ontology xmlns="http://www.w3.org/2002/07/owl#"
     xml:base="urn:absolute:/untitled-ontology-15"
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:xml="http://www.w3.org/XML/1998/namespace"
     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
     ontologyIRI="urn:absolute:/untitled-ontology-15">
    <Prefix name="" IRI="urn:absolute:/untitled-ontology-15#"/>
    <Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#"/>
    <Prefix name="rdf" IRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
    <Prefix name="xml" IRI="http://www.w3.org/XML/1998/namespace"/>
    <Prefix name="xsd" IRI="http://www.w3.org/2001/XMLSchema#"/>
    <Prefix name="rdfs" IRI="http://www.w3.org/2000/01/rdf-schema#"/>
    <Declaration>
        <Class IRI="#BalanceSheetData"/>
    </Declaration>
    <Declaration>
        <Class IRI="#BayesianNetwork"/>
    </Declaration>
    <Declaration>
        <Class IRI="#CAGRvsInflation"/>
    </Declaration>
    <Declaration>
        <Class IRI="#CalculationFormula"/>
    </Declaration>
    <Declaration>
        <Class IRI="#CashAndCashEquivalents"/>
    </Declaration>
    <Declaration>
        <Class IRI="#CashFlowStatementData"/>
    </Declaration>
    <Declaration>
        <Class IRI="#ClassificationScheme"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Classifier"/>
    </Declaration>
    <Declaration>
        <Class IRI="#ConsensusForecasts"/>
    </Declaration>
    <Declaration>
        <Class IRI="#CostOfEquity"/>
    </Declaration>
    <Declaration>
        <Class IRI="#DebtToEquity"/>
```

```xml
</Declaration>
<Declaration>
   <Class IRI="#EarningsPerShare"/>
</Declaration>
<Declaration>
   <Class IRI="#EconomicData"/>
</Declaration>
<Declaration>
   <Class IRI="#EliminationFactor"/>
</Declaration>
<Declaration>
   <Class IRI="#EquityStatementData"/>
</Declaration>
<Declaration>
   <Class IRI="#Factor"/>
</Declaration>
<Declaration>
   <Class IRI="#FinancialMarket"/>
</Declaration>
<Declaration>
   <Class IRI="#FinancialRisk"/>
</Declaration>
<Declaration>
   <Class IRI="#FinancialStatementData"/>
</Declaration>
<Declaration>
   <Class IRI="#FormalOrganisation"/>
</Declaration>
<Declaration>
   <Class IRI="#Formula"/>
</Declaration>
<Declaration>
   <Class IRI="#ForwardEarnings"/>
</Declaration>
<Declaration>
   <Class IRI="#ForwardEarningsGrowthRate"/>
</Declaration>
<Declaration>
   <Class IRI="#ForwardPriceToEarnings"/>
</Declaration>
<Declaration>
   <Class IRI="#FundamentalData"/>
</Declaration>
<Declaration>
   <Class IRI="#Growth"/>
</Declaration>
<Declaration>
   <Class IRI="#HistoricalEarningsGrowthRate"/>
</Declaration>
<Declaration>
   <Class IRI="#HistoricalPriceToEarnings"/>
```

```
</Declaration>
<Declaration>
  <Class IRI="#IncomeStatementData"/>
</Declaration>
<Declaration>
  <Class IRI="#IndustryClassifier"/>
</Declaration>
<Declaration>
  <Class IRI="#IndustrySectorClassificationScheme"/>
</Declaration>
<Declaration>
  <Class IRI="#Inflation"/>
</Declaration>
<Declaration>
  <Class IRI="#IntrinsicValueDCF"/>
</Declaration>
<Declaration>
  <Class IRI="#IntrinsicValueDDM"/>
</Declaration>
<Declaration>
  <Class IRI="#InvestmentAsset"/>
</Declaration>
<Declaration>
  <Class IRI="#IssuedEquity"/>
</Declaration>
<Declaration>
  <Class IRI="#MarketCapitalisation"/>
</Declaration>
<Declaration>
  <Class IRI="#MarketData"/>
</Declaration>
<Declaration>
  <Class IRI="#MarketRateOfReturn"/>
</Declaration>
<Declaration>
  <Class IRI="#ModelData"/>
</Declaration>
<Declaration>
  <Class IRI="#Multiples"/>
</Declaration>
<Declaration>
  <Class IRI="#NegativeEarnings"/>
</Declaration>
<Declaration>
  <Class IRI="#NegativeShareholdersEquity"/>
</Declaration>
<Declaration>
  <Class IRI="#NetIncome"/>
</Declaration>
<Declaration>
  <Class IRI="#PreferenceFactor"/>
```

```
</Declaration>
<Declaration>
    <Class IRI="#PresentDiscountedValue"/>
</Declaration>
<Declaration>
    <Class IRI="#PriceToEarningsRelative"/>
</Declaration>
<Declaration>
    <Class IRI="#Profitability"/>
</Declaration>
<Declaration>
    <Class IRI="#QualityFactor"/>
</Declaration>
<Declaration>
    <Class IRI="#ROEvsCOE"/>
</Declaration>
<Declaration>
    <Class IRI="#Ratios"/>
</Declaration>
<Declaration>
    <Class IRI="#RelativeDebtToEquity"/>
</Declaration>
<Declaration>
    <Class IRI="#ReturnOnEquity"/>
</Declaration>
<Declaration>
    <Class IRI="#RiskFreeRateOfReturn"/>
</Declaration>
<Declaration>
    <Class IRI="#SectorClassifier"/>
</Declaration>
<Declaration>
    <Class IRI="#Share"/>
</Declaration>
<Declaration>
    <Class IRI="#ShareBeta"/>
</Declaration>
<Declaration>
    <Class IRI="#SharePrice"/>
</Declaration>
<Declaration>
    <Class IRI="#ShareholdersEquity"/>
</Declaration>
<Declaration>
    <Class IRI="#SharesInIssue"/>
</Declaration>
<Declaration>
    <Class IRI="#StockMarket"/>
</Declaration>
<Declaration>
    <Class IRI="#StockMarketIndex"/>
```

```
</Declaration>
<Declaration>
   <Class IRI="#SubsectorClassifier"/>
</Declaration>
<Declaration>
   <Class IRI="#SupersectorClassifier"/>
</Declaration>
<Declaration>
   <Class IRI="#SystematicRisk"/>
</Declaration>
<Declaration>
   <Class IRI="#Threshold"/>
</Declaration>
<Declaration>
   <Class IRI="#ThresholdBeta"/>
</Declaration>
<Declaration>
   <Class IRI="#TimeSeriesData"/>
</Declaration>
<Declaration>
   <Class IRI="#TotalAssets"/>
</Declaration>
<Declaration>
   <Class IRI="#TotalEquity"/>
</Declaration>
<Declaration>
   <Class IRI="#TotalLiabilities"/>
</Declaration>
<Declaration>
   <Class IRI="#ValueFactor"/>
</Declaration>
<Declaration>
   <Class IRI="#VariableParameter"/>
</Declaration>
<Declaration>
   <NamedIndividual IRI="#%10"/>
</Declaration>
<Declaration>
   <NamedIndividual IRI="#%20"/>
</Declaration>
<Declaration>
   <NamedIndividual IRI="#Consumer_Price_Inflation_(CPI)"/>
</Declaration>
<Declaration>
   <NamedIndividual IRI="#Producer_Price_Inflation_(PPI)"/>
</Declaration>
<SubClassOf>
   <Class IRI="#BalanceSheetData"/>
   <Class IRI="#FinancialStatementData"/>
</SubClassOf>
<SubClassOf>
```

```
    <Class IRI="#BayesianNetwork"/>
    <Class IRI="#ClassificationScheme"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#CAGRvsInflation"/>
    <Class IRI="#Growth"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#CalculationFormula"/>
    <Class IRI="#Formula"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#CalculationFormula"/>
    <Class IRI="#PresentDiscountedValue"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#CashAndCashEquivalents"/>
    <Class IRI="#BalanceSheetData"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#CashFlowStatementData"/>
    <Class IRI="#FinancialStatementData"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#ConsensusForecasts"/>
    <Class IRI="#MarketData"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#CostOfEquity"/>
    <Class IRI="#CalculationFormula"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#DebtToEquity"/>
    <Class IRI="#CalculationFormula"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#EarningsPerShare"/>
    <Class IRI="#IncomeStatementData"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#EconomicData"/>
    <Class IRI="#ModelData"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#EliminationFactor"/>
    <Class IRI="#Factor"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#EquityStatementData"/>
    <Class IRI="#FinancialStatementData"/>
</SubClassOf>
```

```xml
<SubClassOf>
    <Class IRI="#FinancialRisk"/>
    <Class IRI="#QualityFactor"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#FinancialStatementData"/>
    <Class IRI="#FundamentalData"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#ForwardEarnings"/>
    <Class IRI="#ConsensusForecasts"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#ForwardEarningsGrowthRate"/>
    <Class IRI="#CalculationFormula"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#ForwardPriceToEarnings"/>
    <Class IRI="#CalculationFormula"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#ForwardPriceToEarnings"/>
    <Class IRI="#Multiples"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#Growth"/>
    <Class IRI="#QualityFactor"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#HistoricalEarningsGrowthRate"/>
    <Class IRI="#CalculationFormula"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#HistoricalPriceToEarnings"/>
    <Class IRI="#CalculationFormula"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#HistoricalPriceToEarnings"/>
    <Class IRI="#Multiples"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#IncomeStatementData"/>
    <Class IRI="#FinancialStatementData"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#IndustryClassifier"/>
    <Class IRI="#Classifier"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#IndustrySectorClassificationScheme"/>
    <Class IRI="#ClassificationScheme"/>
```

```xml
</SubClassOf>
<SubClassOf>
   <Class IRI="#Inflation"/>
   <Class IRI="#EconomicData"/>
</SubClassOf>
<SubClassOf>
   <Class IRI="#IntrinsicValueDCF"/>
   <Class IRI="#CalculationFormula"/>
</SubClassOf>
<SubClassOf>
   <Class IRI="#IntrinsicValueDDM"/>
   <Class IRI="#CalculationFormula"/>
</SubClassOf>
<SubClassOf>
   <Class IRI="#IntrinsicValueDDM"/>
   <Class IRI="#PresentDiscountedValue"/>
</SubClassOf>
<SubClassOf>
   <Class IRI="#MarketCapitalisation"/>
   <Class IRI="#CalculationFormula"/>
</SubClassOf>
<SubClassOf>
   <Class IRI="#MarketData"/>
   <Class IRI="#FundamentalData"/>
</SubClassOf>
<SubClassOf>
   <Class IRI="#MarketRateOfReturn"/>
   <Class IRI="#VariableParameter"/>
</SubClassOf>
<SubClassOf>
   <Class IRI="#Multiples"/>
   <Class IRI="#ValueFactor"/>
</SubClassOf>
<SubClassOf>
   <Class IRI="#NegativeEarnings"/>
   <Class IRI="#EliminationFactor"/>
</SubClassOf>
<SubClassOf>
   <Class IRI="#NegativeShareholdersEquity"/>
   <Class IRI="#EliminationFactor"/>
</SubClassOf>
<SubClassOf>
   <Class IRI="#NetIncome"/>
   <Class IRI="#IncomeStatementData"/>
</SubClassOf>
<SubClassOf>
   <Class IRI="#PreferenceFactor"/>
   <Class IRI="#Factor"/>
</SubClassOf>
<SubClassOf>
   <Class IRI="#PresentDiscountedValue"/>
```

```xml
    <Class IRI="#ValueFactor"/>
  </SubClassOf>
  <SubClassOf>
    <Class IRI="#PriceToEarningsRelative"/>
    <Class IRI="#CalculationFormula"/>
  </SubClassOf>
  <SubClassOf>
    <Class IRI="#PriceToEarningsRelative"/>
    <Class IRI="#Ratios"/>
  </SubClassOf>
  <SubClassOf>
    <Class IRI="#Profitability"/>
    <Class IRI="#QualityFactor"/>
  </SubClassOf>
  <SubClassOf>
    <Class IRI="#QualityFactor"/>
    <Class IRI="#Factor"/>
  </SubClassOf>
  <SubClassOf>
    <Class IRI="#ROEvsCOE"/>
    <Class IRI="#Profitability"/>
  </SubClassOf>
  <SubClassOf>
    <Class IRI="#Ratios"/>
    <Class IRI="#ValueFactor"/>
  </SubClassOf>
  <SubClassOf>
    <Class IRI="#RelativeDebtToEquity"/>
    <Class IRI="#CalculationFormula"/>
  </SubClassOf>
  <SubClassOf>
    <Class IRI="#RelativeDebtToEquity"/>
    <Class IRI="#FinancialRisk"/>
  </SubClassOf>
  <SubClassOf>
    <Class IRI="#ReturnOnEquity"/>
    <Class IRI="#CalculationFormula"/>
  </SubClassOf>
  <SubClassOf>
    <Class IRI="#RiskFreeRateOfReturn"/>
    <Class IRI="#EconomicData"/>
  </SubClassOf>
  <SubClassOf>
    <Class IRI="#SectorClassifier"/>
    <Class IRI="#SupersectorClassifier"/>
  </SubClassOf>
  <SubClassOf>
    <Class IRI="#Share"/>
    <Class IRI="#InvestmentAsset"/>
  </SubClassOf>
  <SubClassOf>
```

```
    <Class IRI="#Share"/>
    <Class IRI="#IssuedEquity"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#ShareBeta"/>
    <Class IRI="#TimeSeriesData"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#SharePrice"/>
    <Class IRI="#TimeSeriesData"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#ShareholdersEquity"/>
    <Class IRI="#TotalEquity"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#SharesInIssue"/>
    <Class IRI="#IncomeStatementData"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#StockMarket"/>
    <Class IRI="#FinancialMarket"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#StockMarketIndex"/>
    <Class IRI="#InvestmentAsset"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#SubsectorClassifier"/>
    <Class IRI="#SectorClassifier"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#SupersectorClassifier"/>
    <Class IRI="#IndustryClassifier"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#SystematicRisk"/>
    <Class IRI="#PreferenceFactor"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#Threshold"/>
    <Class IRI="#VariableParameter"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#ThresholdBeta"/>
    <Class IRI="#SystematicRisk"/>
</SubClassOf>
<SubClassOf>
    <Class IRI="#TimeSeriesData"/>
    <Class IRI="#MarketData"/>
</SubClassOf>
```

```xml
    <SubClassOf>
        <Class IRI="#TotalAssets"/>
        <Class IRI="#BalanceSheetData"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#TotalEquity"/>
        <Class IRI="#BalanceSheetData"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#TotalLiabilities"/>
        <Class IRI="#BalanceSheetData"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#ValueFactor"/>
        <Class IRI="#Factor"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#VariableParameter"/>
        <Class IRI="#ModelData"/>
    </SubClassOf>
    <ClassAssertion>
        <Class IRI="#Threshold"/>
        <NamedIndividual IRI="#%10"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="#Inflation"/>
        <NamedIndividual IRI="#Consumer_Price_Inflation_(CPI)"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="#Inflation"/>
        <NamedIndividual IRI="#Producer_Price_Inflation_(PPI)"/>
    </ClassAssertion>
</Ontology>



<!-- Generated by the OWL API (version 4.5.9.2019-02-01T07:24:44Z) https://github.com/owlcs/owlapi -->
```