Network for Studies on Pensions, Aging and Retirement

Applied Data Science in the Pension Industry: A Survey and Outlook

Onaopepo Adekunle
Michel Dumontier
Arno Riedl

DESIGN PAPER 183

NETSPAR INDUSTRY SERIES

DESIGN PAPERS are part of the **refereed Industry Paper Series**, which are refereed by the Netspar Editorial Board. Design Papers discuss the design of a component of a pension system or product. A Netspar Design Paper analyzes the objective of a component and the possibilities for improving its efficacy. These papers are easily accessible for industry specialists who are responsible for designing the component being discussed. Authors are allowed to give their personal opinion in a separate section. Design Papers are presented for discussion at Netspar events. Representatives of academic and private sector partners, are invited to these events. Design Papers are published at the Netspar website.

Colophon
Netspar Design Paper 183, September 2021

Design
B-more Design

Lay-out
Bladvulling, Tilburg

Editors
Frans Kooymans, Frans Kooymans-Text and Translation
Netspar

CONTENTS

Affiliations

Onaopepo Adekunle – Institute of Data Science, Maastricht University

Michel Dumontier – Institute of Data Science, Maastricht University

Arno Riedl – School of Business and Economics, Maastricht University

Abstract

The pension industry, much like the rest of the financial industry, is increasingly adopting data science and artificial intelligence-based solutions. Applications range from leaner and faster operations ("doing the same thing better") to completely new value propositions. However, the available literature suggests that the pension industry appears to be relatively conservative and cautious when it comes to adopting new and dynamically changing machine learning (ML) techniques. The black box nature of most ML techniques also appears to contribute to the skepticism of the pension sector. Hence, there seems to be a gap between the potential applications of data science solutions proposed by researchers and their application in the pension industry. This article provides (i) a review of what has been reported in the data science literature, (ii) a taxonomy of ML techniques that can be applied for challenges in the pension industry, and (iii) a categorization of the different aspects of the pension industry that are covered in state-of-the-art applied data science.

We surveyed 25 papers and presentations on the application of data science in the pension industry and highlight the major machine learning techniques that were used and their applicability in the pension sector. These techniques are concisely introduced to provide a basis for stakeholders to gain an understanding of their potential applicability to tackle challenges in the pension industry. Based on the existing research, three areas of the pension industry are identified as most relevant for the application of machine learning techniques: customer focus, organizational process optimization, and personnel optimization. Open issues and further opportunities regarding the application of data science in the pension sector are discussed.

We surveyed the existing body of literature to summarize how data science is being currently leveraged to deal with issues related to pensions. Prominent developments appear along the fronts of prediction and chatbot development. Our analysis suggests that there remains ample room in the pension industry to explore the use of other machine learning and data mining methodologies, such as clustering, natural language processing, and reinforcement learning. This includes gleaning insights from unconventional sources such as social media activity, and developing new customer-focused and business development applications.

Samenvatting
Toegepaste datawetenschap in de pensioensector: een overzicht en vooruitblik

De pensioensector maakt in toenemende mate gebruik van oplossingen die gebaseerd zijn op datawetenschap. Dit artikel geeft (i) een overzicht van wat hierover in de datawetenschapsliteratuur bekend is, (ii) een taxonomie van zelflerende technieken (machine learning)– die toegepast kunnen worden in de pensioensector en (iii) een categorisering van de verschillende aspecten van de pensioensector waarvoor de toegepaste state-of-the-art datawetenschap oplossingen biedt.

We hebben 25 artikelen onderzocht naar toepassingen van datawetenschap in de pensioensector. We belichten de belangrijkste machine learning-technieken die zijn gebruikt en bespreken hun inzetbaarheid in de pensioensector. Deze technieken worden kort voorgesteld en op basis van het bestaande onderzoek worden drie gebieden van de pensioensector geïdentificeerd die het meest kunnen profiteren van de toepassing van machine learning-technieken: klantgerichtheid, optimalisering van organisatorische processen en optimalisering van personeel. Openstaande vraagstukken en verdere mogelijke toepassingen van datawetenschap in de pensioensector worden besproken.

Uit onze analyse blijkt dat in de pensioensector nog veel ruimte is voor onderzoek naar de inzet van machine learning en datamining, zoals voor clustering, natuurlijke taalverwerking en ondersteund leren, om inzichten te verkrijgen vanuit onconventionele bronnen, bijvoorbeeld activiteiten op sociale media, en voor de ontwikkeling van nieuwe toepassingen op het gebied van klantgerichtheid en bedrijfsontwikkeling.

1. Introduction

It has been suggested that the pensions industry lags behind other financial sectors in its adoption of data science and artificial intelligence (AI) technologies, including machine learning and natural language processing[1], largely owing to a lack of expertise and focused strategic investment. Data science and AI offer new approaches in data analysis, optimization, and pattern recognition that could be applied to address long-standing pension challenges, and that could be the vehicles towards new commercial possibilities as well as support for individual decisions (Kaufmann et al. 2018). Automated processing of large data volumes may lead to more reliable and accurate predictive models (Talwar & Kumar 2013). The reliance of AI on data also requires a fundamental shift in corporate culture towards improvement in data management practices (McAfee & Brynjolfsson 2012).

The purpose of this article is to give an overview of current developments in the application of data science and artificial intelligence in the pension industry and related financial services.

The main contributions include:

1. An overview of how data science and AI are being applied to pension research and systems.
2. An examination of machine learning techniques that can be adopted for pension-based scenarios.
3. A discussion of open issues relating to the adoption of data science in the pension industry.

We surveyed 25 papers and presentations on the application of data science in the pension industry and highlight the major machine learning techniques that were presented in these papers as well as their applications. The main aspects of these techniques are introduced in this article and tagged with representative reference, presented in Table 1 below.

The rest of this article is organized as follows. The broad application of data science in the financial services industry is summarized in Section 2. Section 3 explores the taxonomy of applicable machine learning algorithms for the pensions industry. In Section 4 a literature review of data science research applied to the pension industry is provided. Section 5 highlights some open issues regarding the adoption of data science in the pension industry, and Section 6 provides concluding remarks.

1 *https://www.pensionsage.com/pa/images/PA_Oct_2018_AI.pdf*

2. Data Science in Financial Services

Machine learning has achieved remarkable success in many domains. Its potential has increased due to developments in new technologies such as neural networks. Examples of applicable data science solutions include advanced chatbots, identity verification in client onboarding (i.e, welcoming new clients), transaction data analysis, fraud detection in claims management, pricing in bond trading, price differentiation in car insurance, automated analysis of legal documents, customer relation management, risk management, portfolio management, trading execution, and investment operations. The pension industry still has a lot of room to capitalize on the benefits of machine learning and data science. This paper highlights the ground already covered by researchers, by reviewing the literature on the data science approaches that have been applied in the pension domain. As such it illustrates to industry professionals what can be achieved and provides an overview for interested researchers.

Many machine learning algorithms operate in real time and update as new data are entered, which fits well with challenges in financial services such as portfolio management and stock market prediction. In other domains, platforms such as Amazon and Netflix predict in real time the preferences of their customers based on previous choices (Bell et al. 2008); this could also be applied to customers in the pensions industry. Models for prediction of weather changes also update continuously online as environmental conditions change (Alavi et al. 2016); they can be applied for pricing in bond trading and the like. The accuracy of these models is the result of the training process and automation that are part of machine learning. Within the financial sector, machine learning algorithms such as decision tree and neural networks have been used to detect credit card fraud. These algorithms help companies to minimize the losses from such financial crime. In addition, when coupled with a real time system, this method can quickly inform clients and rapidly resolve problems (Bolton & Hand 2001, Delamaire et al. 2009, Juszczak et al. 2008, Pozzolo et al. 2014). Studies in the insurance sector have used artificial neural networks to evaluate the financial capability of insurance companies and to predict insolvency (Olaniyi et al. 2012). Other studies have used feedforward neural networks with the back-propagation algorithm to build decision models for five insurance categories including life, annuity, health, accident, and investment-oriented insurance (Lin et al. 2008). Machine learning algorithms also have been used to analyze the quality of mortality models (Deprez et al. 2017). Bacham & Zhao (2017) analyzed the performance of a set of machine learning methods in assessing credit risk of small and medium-sized borrowers, with Moody's

Analytics RiskCalc model serving as the benchmark model. We elaborate on how to use data science in the next section by providing an in-depth description of the various learning techniques and providing examples of where they can be applied.

Econometrics and machine learning models

Econometrics and ML models both apply statistical methods to make inferences on data. However, according to Frisch (1933) in the very first issue of *Econometrica*, econometrics is about "economic theory in its relation to statistics and mathematics". The main purpose of econometrics is not to predict but to quantify an economic phenomenon. Thus, econometric models often test hypotheses based on economic theory or game theory and make assumptions about the data investigated. On the contrary, data science and machine learning are driven by data rather than theory, not making any assumptions on the data. Thus, the most important difference between classical econometric models and ML is that the former focus on estimation and the testing of hypotheses, often in smaller samples, whereas the latter focuses on the best functional approximation, often in huge samples. Moreover, econometric modeling mostly focuses on parametric methods making distributional assumptions, ML more often (but not exclusively) applies non-parametric distribution-free methods.

3. Learning Techniques

Machine learning is a subfield of computer science. It is a type of Artificial Intelligence (AI) that uses historical data as input to predict new output values. It uses machines to enable learning with new information without the need to rewrite the program. Machine learning evolved from pattern recognition and computational learning theory. Here we elaborate on some essential concepts of machine learning as well as on frequently applied machine learning algorithms for smart data analysis.

A learning algorithm takes a set of samples as input. This is called a training set. In general, there are four main learning categories: supervised, unsupervised, semi-supervised, and reinforcement learning (Bishop, 2006; Barber, 2012; Murphy, 2012). In supervised learning, the training set consists of samples of input variables together with a target variable of interest, also known as labels. In unsupervised learning, no labels are required for the training set. Semi-supervised learning deals with the case of having a small number of labels for some of the samples in the training set while most labels are missing.

Reinforcement learning (RL) deals with the problem of learning the appropriate action or sequence of actions to be taken for a given situation to maximize payoff. According to Snow (2019), reinforcement learning in finance comprises the use of an agent that learns how to take actions in an environment to maximize some notion of cumulative reward. The agent exists in a predefined environment and receives as input the current state and is then asked to take an action to receive a reward, the information of which can be used to identify the next optimal action. The benefit of reinforcement learning algorithms is that the final objective function can be the realized/unrealized profit and loss, but also values like the Sharpe Ratio, maximum drawdown, and value at risk measures. RL allows for end-to-end optimization[2] on what maximizes rewards. The RL algorithm directly learns a policy. For instance, reinforcement learning answers the question, "Should I buy the asset today?", whereas supervised learning answers the question, "Will the price of the asset increase tomorrow?". Thus, reinforcement learning algorithms lend themselves well for developing trading strategies.

This paper focuses on supervised and unsupervised learning, since they have been and continue to be widely applied in the financial services industry. The highlighted machine learning models are the most prominently used data science models in the

2 End-to-end optimization entails the optimization process over a system or service from beginning to end until the delivery of a complete functional solution.

pension industry. The objective of supervised learning is to learn how to predict the appropriate output variable for given input variables. Applications where the target labels consist of a finite number of discrete categories are known as classification tasks. If the dependent variables are text-based, the aim is to determine whether the text leans towards positive or negative classification; this is called sentiment analysis. Cases where the target labels are comprised of one or more continuous variables are known as regression tasks (Bengio et al. 2016).

Defining the objective of unsupervised learning is more difficult. One of the major objectives is to identify viable clusters of similar samples within the input data; this is known as clustering. The objective may also be to discover a useful internal representation of the input data by pre-processing the original input variable in order to transfer it into a new variable space. This pre-processing stage can significantly improve the result of the subsequent machine learning algorithm. Bishop (2006) aptly calls it feature extraction.

3.1 Supervised Learning Techniques

Linear regression

Linear regression is a method that is used to model the linear relationship between a dependent variable (target) and one or more independent variables (predictors). It is based on the statistical ordinary least squares (OLS) method, where the model is fit such that the sum-of-squares of differences of observed and predicted values is minimized. The error function is minimized as an optimization function to estimate the coefficients (*bi*). To avoid overfitting, this can be further developed to include a regularization term in the optimization function. Overfitting happens when the model captures the noise in the training dataset. By noise we mean the data points

$$\textit{observed data} \rightarrow \quad y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p + \varepsilon$$

$$\textit{predicted data} \rightarrow \quad y' = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p$$

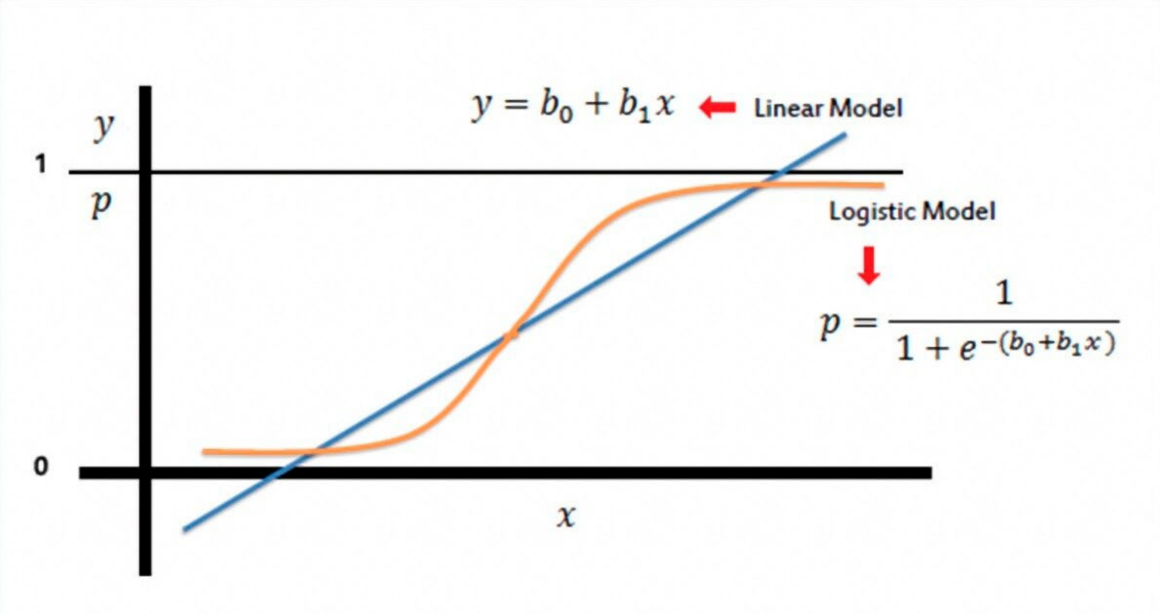
$$\textit{error} \rightarrow \quad \varepsilon = y - y'$$

$$\textit{minimize} \sum (y - y')^2 \qquad \text{OLS}$$

$$\textit{minimize} \sum (y - y')^2 + \lambda \sum |b| \qquad \text{Lasso}$$

$$\textit{minimize} \sum (y - y')^2 + \lambda \sum b^2 \qquad \text{Ridge}$$

Figure 1 Logistic regression (Sayad, 2012)



that do not really represent the true properties of the data, but instead chance. A regularizer shrinks the coefficient estimates towards zero to mitigate overfitting and, thereby, low accuracy of the model. When the regularizer is based on the so-called L1-normalization of the coefficients, it is referred to as Lasso regression. When the regularizer is based on the so-called L2-normalization of the coefficients, it is called a Ridge regression.

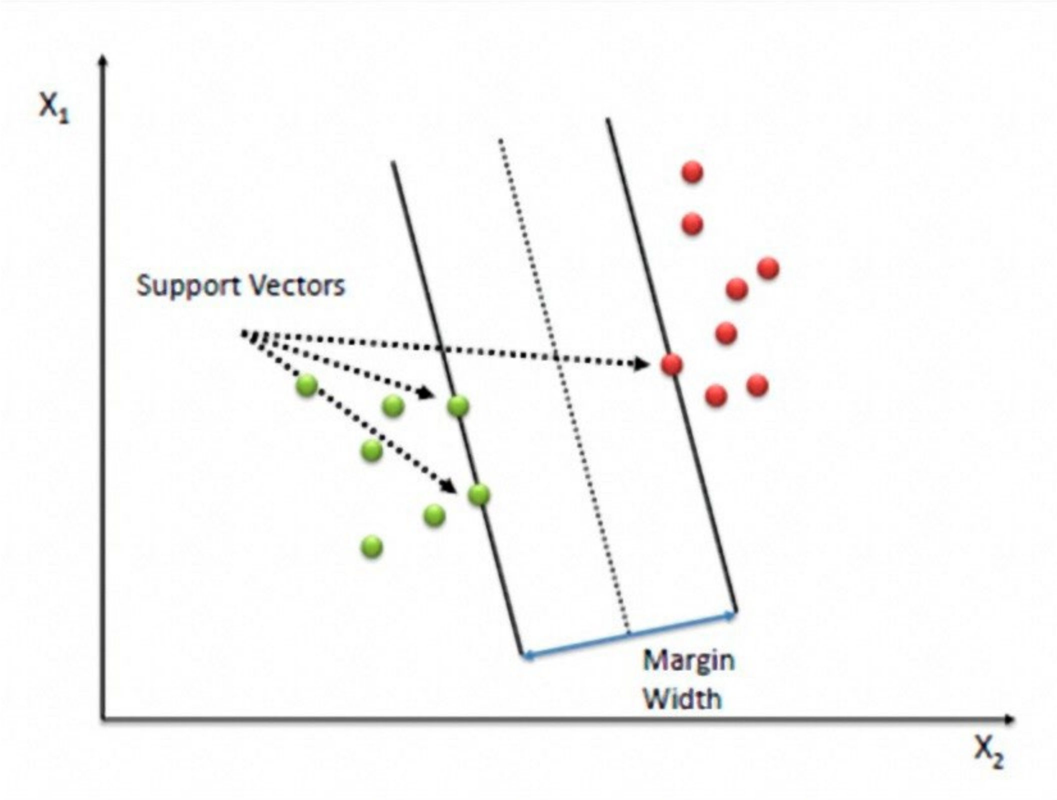
Logistic regression

A logistic regression predicts the probability of an outcome that can only have two values (i.e. a binary variable). A logistic regression produces a logistic curve as depicted in Figure 1, which is limited to values between 0 and 1. The curve is constructed using the natural logarithm of the "odds" of the target variable, rather than the probability (Hosmer Jr et al., 2013). Moreover, the predictors do not have to be normally distributed or have equal variance in each group. Figure 1, taken from Sayad (2012), shows an example of a logistic curve in comparison with a linear regression.

Support vector machines

Support Vector Machines (SVM) perform classification by representing the predictor variable space in a higher dimensional variable space. This is achieved through mathematical transformation of the variables and finding the separating hyperplane that maximizes the margin between the two classes. The vectors (cases) that define the hyperplane are the support vectors (Cortes & Vapnik, 1995). Figure 2, taken from Sayad

Figure 2 Support vector machine (Sayad, 2012)

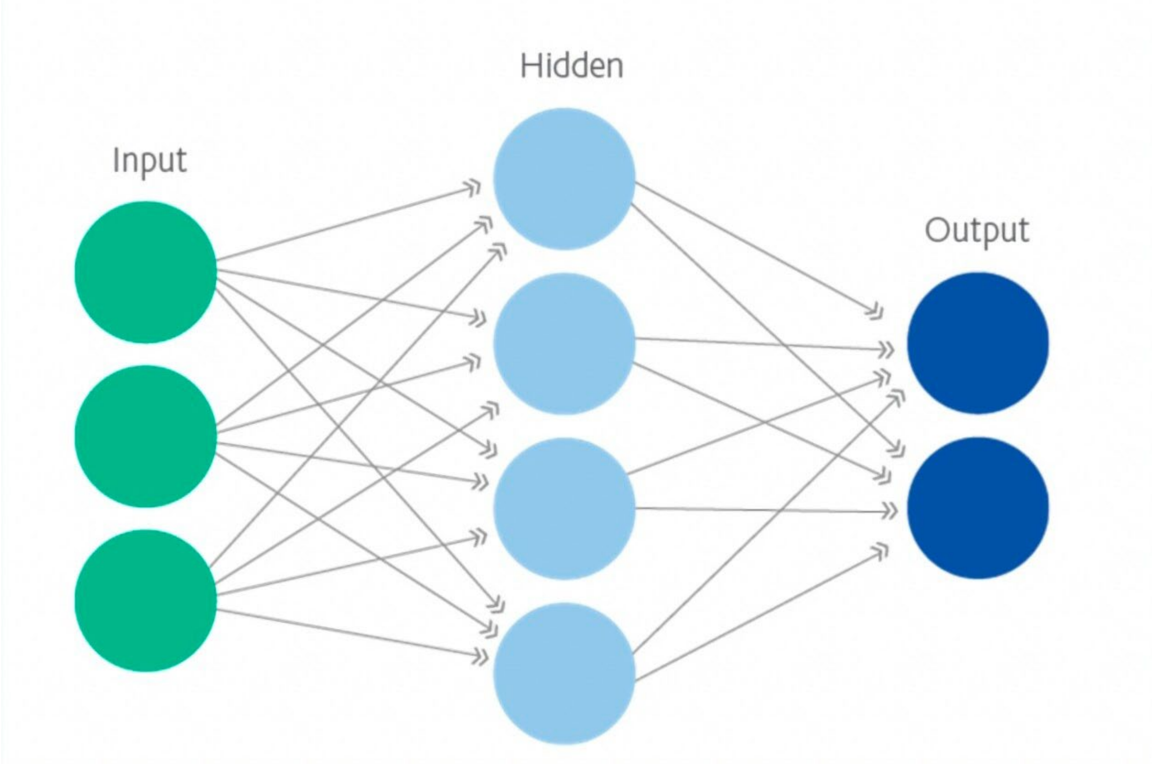


(2012), shows an example of support vectors on a linearly separable variable space for hypothetical observation data. In the figure the dotted line without arrowhead is the line that linearly separates different classes of the data points. The line is computed such that it maximizes the minimum distance between data points; these data points are called support vectors (see Figure 2). SVM differs from the other classification algorithms in that it chooses the decision boundary that maximizes the distance from the nearest data points of all classes. SVM does not merely find a random decision boundary; it finds the optimal decision boundary.

Artificial neural networks

An artificial neural network (ANN) is a mathematical simulation of a biological neural network. Its simple form is shown in Figure 3 (see Bacham & Zhao, 2017). In this example, there are three input values and two output values. An ANN consists of a network of artificial neurons (also known as "nodes"). These nodes are connected to each other, and to these connections values are assigned reflecting their strength: inhibition (maximum −1.0) or excitation (maximum +1.0). If the absolute value of the connection is high, then it indicates a strong connection. Different transformations

Figure 3 Artificial Neural Network (Bacham & Zhao, 2017)



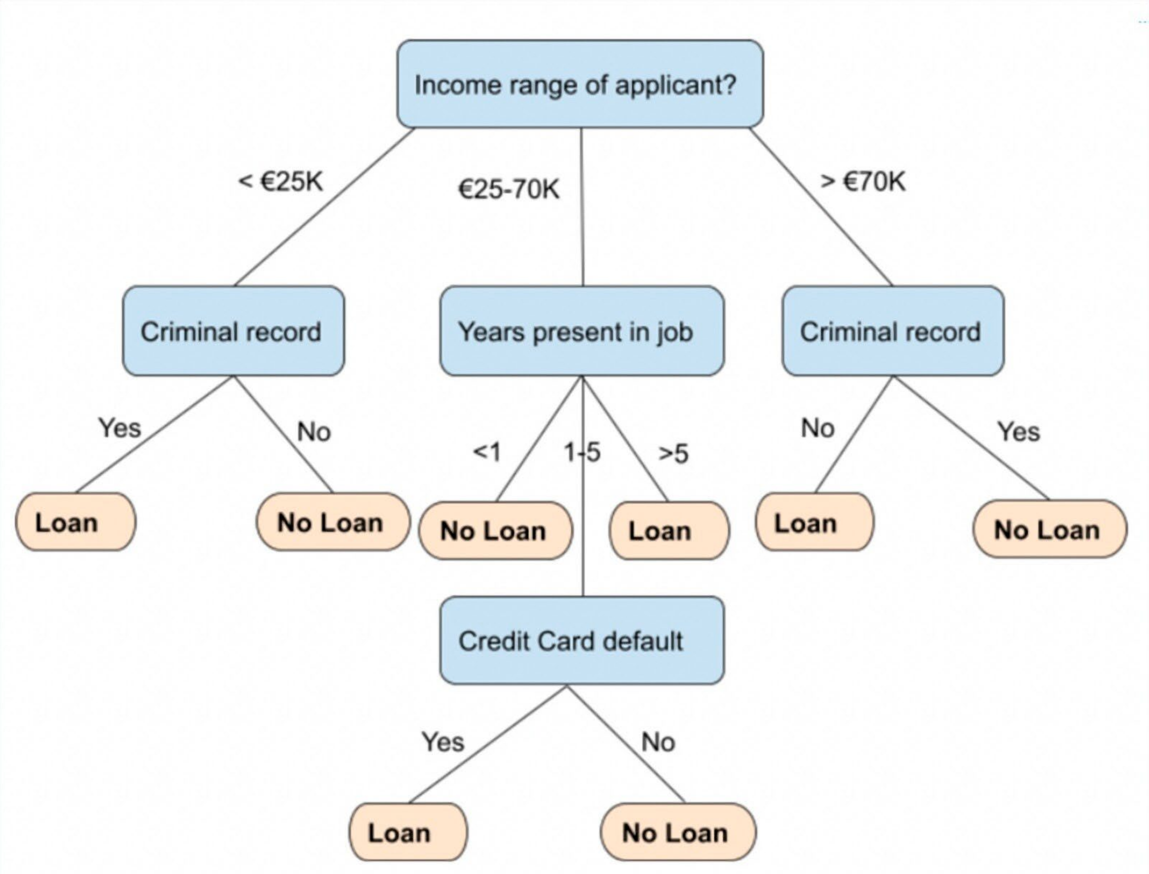
link the input values to a hidden layer, and the hidden layer to the output values. ANNs can easily handle non-linear and interactive effects of explanatory variables due to the presence of many hidden layers and neurons.

Decision trees and random forest

In a decision tree, an input is entered at the top; as it traverses down the tree, the data are bucketed into ever smaller subsets. A decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is developed incrementally. The final result is a tree with decision nodes and leaf nodes. In the hypothetical example shown in Figure 4, the tree determines whether to provide a loan to an individual based on four variables: income range of the person, years present in the job, credit card default status, and existence of a criminal record.

A decision tree is the most basic unit of the random forest. Random forests combine decision tree predictors, such that each tree depends on the values of a random observation list sampled independently, and with the same distribution. Thus, the random forest approach combines the predictions of many trees, and the final decision is based on the average of the outputs of the underlying independent decision trees.

Figure 4 Decision tree for loan procurement



3.2 Unsupervised Learning Techniques

K-means clustering

The objective of K-means clustering is to cluster the unlabeled data set into K clusters (groups), where data points belonging to the same cluster must have some similarities. As such, the distance between data points is the measure of similarity. Therefore, K-means clustering seeks to find a set of K cluster centers, such that the distances between data points and their nearest center are minimized (Coates & Ng, 2012).

An example of a simplified algorithm:

1) Clusters the data into K groups, where K is predefined.
2) Select K data points at random as cluster centers.
3) Assign objects to their closest cluster center according to the Euclidean distance function.
4) Calculate the centroid or mean of all objects in each cluster.
5) Repeat steps 2, 3, and 4 until the same points are assigned to each cluster in consecutive rounds.

In practice, K-means is a very fast and highly scalable algorithm. Moreover, there is an online stochastic version of K-means (Jumutc et al., 2015) that enables it to scale to big data application. However, this approach also has limitations because it uses the Euclidean distance as measure of similarity. This restricts the types of data variables that can be considered, and cluster centers are not robust against outliers. Additionally, the K-means algorithm assigns each data point to only one of the clusters, which may lead to inappropriate clusters in some cases (Likas et al., 2003).

For proper data science and analysis, it is necessary to determine which task should be accomplished out of the various possibilities, ranging from understanding the internal structure of the data to finding unusual data points and to prediction of values and categories. The current body of literature focuses on prediction tasks, which leaves room to conduct the other types of tasks mentioned. Here we have limited the scope of our review to supervised and unsupervised learning techniques, highlighting the techniques applied most widely in the pensions industry.

To discover the structure of unlabeled data, clustering algorithms can provide the most appropriate tools. K-means clustering, described above, is the best known and most frequently applied clustering algorithm. It can handle large data volumes with a broad range of data types.

The linear regression, random forest, and SVM methods described above are the three most frequently applied algorithms to predict values and classify sequenced data. The objective of the models applied in these algorithms is to process and train data at high velocity.

The most accessible and interpretable way for predicting the categories of data is logistic regression. Neural networks are also suitable learning models for function approximation problems. Moreover, because the volume of pension data is often huge and varied, it requires extensive training; an appropriate solution for this is a multi-class neural network. SVM is another popular classification algorithm, which is capable of handling massive amounts of data and classifying their different types. Because SVM can handle a high volume and a variety of data types, it is commonly applied in data processing algorithms. Caruana & Niculescu-Mizil (2006) provide comparisons of the data science techniques covered in this review.

4. Data Science in Pensions

One way of applying data science in portfolio management is to conduct detailed market simulations. By analyzing swathes of data, algorithms can exploit hundreds of investing patterns all at once, predict investor behavior, and help provide more targeted results for the client. The rise of big data means that fund managers have access to more information than ever before when making decisions for their clients. Applying ML models to analyze the data opens up new opportunities and, as a result, enhances the investment process.

Table 1 provides an overview, based on the papers reviewed, of the machine learning algorithms used for pension data analysis, the cases where it was applied, the type of machine learning tasks, and references where the respective techniques were applied.

We identified three areas where machine learning can lead to better decision making in the pensions industry:

1) Customer-focused approach
2) Organizational process optimization
3) Personnel optimization

Customer-focused approach

Here the focus is on the customers of the pensions industry, in particular retirees. For example, Salazar & Boado-Penas (2019) employed machine learning techniques (logistic regression, SVM, and random forest) to predict early retirement, using data from private pension plans in Mexico. Social and macroeconomic variables were used as predictors for whether an individual retires before the age of 65.

The *Finnish Centre for Pensions* announced that it had taught a machine-learning algorithm to predict retirement on a disability pension, based on socioeconomic, earnings, and benefits data. It achieved a 78% accuracy rate of detecting who is likely to retire on a disability pension[3]. *Mercer Global Australian Business* recently launched the Mercer 'Superbot', which is a financial advice chatbot accessed through Facebook Messenger[4].

Leung et al. (2019) used tweets from *StockTwits* and sentiment analysis from natural language processing (NLP), which is akin to logistic regression to predict stock

3 https://www.actuview.com/predicting-disability-pensions-with-machine-learning-classification-models_cb510e136.html

4 https://www.mercer.com.au/our-thinking/superannuation/meet-superbot-artificial-intelligence-delivering-financial-advice.html

Table 1. Overview of machine learning algorithms for pension data analysis (based on the papers reviewed).

Machine learning technique	Cases	Data processing tasks	Representative references
Logistic regression	Retirement prediction	classification	Salazar & Boado-Penas (2019), Bosworth & Burtless (2010) , Coile & Levine (2011), Blekesaune & Skirbekk (2012), Feldman & Beehr (2011)
Sentiment analysis	Stock return prediction	classification	Leung et al. (2019)
K-means clustering	Pension scheme selection	classification/ recommendation	Dong et al. (2017), Cong et. al (2016), Gough & Sozou (2005)
SVM	Retirement prediction, asset management	classification/ regression	Salazar & Boado-Penas (2019)
Decision tree	Pension fund asset valuation	classification	Aguirre et al. (2019)
Random forest	Portfolio management, asset management	classification/ regression	Salazar & Boado-Penas (2019)
Lasso regression	Defined benefit obligation	regression	Hendriksen (2017), Rapach et al. (2018)
Reinforcement learning	Portfolio management	classification	Cong et al. (2020) Snow (2019)
ANN	pension cost dependency ratio, stock returns prediction, asset management, emotion recognition	classification/ regression	Agnieszka (2018), Avramov et al. (2019), Henkel et al. (2020)

return trends. They found that social-media sentiment positively and significantly predicts future stock returns, and, importantly, that such positive predictability decreases when the number of stocks that users follow increases.

Dong et al. (2017) and Cong et. al (2016) developed a recommendation system from customer attribute information including age and gender. Together with a user interest model based on online interfaces provided by pension service providers such as click-through rate, they developed a recommendation system of pension services to similar users.

Gough & Sozou (2005) studied the variation in behavior and attitudes such as debt tendency with regards to pensions and retirement saving among consumers of financial service products. Using data obtained via a questionnaire, they carried out K-means clustering analysis as the methodology for their study.

Various other studies have been conducted to determine the explanatory factors of retirement decisions. Some studies find that macroeconomic factors such as unemployment rate and stock market performance affect the decision of people to enter retirement early or to delay it (Bosworth & Burtless 2010, Coile & Levine 2011).

Other studies show that personality traits can predict the timing and routes of a person's retirement (Blekesaune & Skirbekk 2012, Feldman & Beehr 2011). In the context of a private company such as an insurance company that offers retirement plans, the retirement decisions are of interest because an early retirement involves fewer contributions to the system, less investment, and therefore lower net earnings. More generally, companies need to predict when and how much of the budgetary provision will be needed to make payments when pensions are claimed.

Organizational process optimization

The goal of process optimization is to significantly improve the core operations of the pension provider such as portfolio management, asset management, and stock forecasting. Snow (2019) highlights various machine learning techniques for portfolio optimization, risk management, and capital management[5]. This author also provides numerous examples that touch on most of the aforementioned machine learning techniques and their financial application, as follows:

Werpachowska (2018) analyzed and forecasted the pension cost dependency ratio for England and Wales from 1991 to 2061, in a bid to evaluate the impact of current state pension reforms and changes in international migration patterns under different Brexit scenarios. She analyzed mortality rate model based on deep learning techniques to account for the volatility in life expectancy, to discover complex patterns in the data as well, and to extrapolate trends. The results show that the recent reforms can effectively stave off the "pension crisis" and bring the system back on sounder fiscal footing.

Aguirre et al. (2019) developed a model that allows one to predict the pension fund of an affiliate in the private pension system by means of a web solution. A boosted decision tree which is similar to the random forest was used to create the model and evaluated using the Pension Fund Administrator (AFP) in Lima (Peru).

Hendriksen (2017) investigated the feasibility of predicting defined benefit obligation (DBO) using a statistical/machine learning approach with some success regarding accuracy of prediction. Salary, age, gender, and the ratio of back service to total service time were explored as explanatory variables, using a Generalized Linear Model.

Otranto & Trudda (2008) used an agglomerative clustering algorithm, based on the distance between time series model of the volatility for different pension funds, applying GARCH models. This provided a classification of the funds based on different

5 This paper provides various examples, which can be found at: *https://ssrn.com/abstract=3420952*

degrees of risk, since similar GARCH models represent similar volatility behavior and similar investment risks.

Avramov et al. (2019) examined empirical evidence of the economic importance and statistical reliability of ANN to predict cross-sectional stock returns and asset management.

Rapach et al. (2018) used lasso regressions to analyze industry return predictability based on the information in lagged industry returns. Controlling for post-selection inference and multiple testing, they found significant in-sample evidence of industry return predictability.

Cong et al. (2020) developed a reinforcement-learning-based portfolio management to directly optimize investors' objectives. This is an alternative that improves upon the traditional two-step portfolio construction.

In Denmark, *PensionDanmark* has automated around 80% of its administrative decisions and aims to increase this further. In the United Kingdom, the Department for Work and Pensions is using AI to crack down on benefits fraud[6].

This overview shows that the area of business optimization uses a diverse set of data science applications, covering many types of machine learning techniques that range from supervised techniques to reinforcement learning.

Personnel optimization

Sasaki et al. (2018) studied the feasibility of using AI for fund manager structure development at the behest of the Japanese Government Pension Investment Fund (GPIF). A joint team, consisting of GPIF and Sony representatives, examined the management structure development and maintenance processes of GPIF. They explored the use of a proof-of-concept prototype system to test the principle of applying ANNs to detect the investment style of managers, based on trading behavior data (trading items, timing, volume, unrealized gain and loss, etc.) collected on a daily basis by GPIF.

Essentially, research that focuses on augmenting the personnel workspace in a pension organization falls in the personnel optimization category. Another example is Henkel et al. (2020), who developed AI-based emotion recognition software that helps service employees in managing customer emotions while interacting with them. The case scenario of focus was two pension funds in the Netherlands that provided access to their centralized call center operations.

6 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/721224/dwp-annual-report-and-accounts-2017-2018.pdf

5. Discussion

In order to exploit the full potential of data science, we identify three major challenges for the pension industry, which are described below. In addition, we briefly discuss the similarities and differences between econometrics models and machine learning models.

Data collection and quality

Because data are the basis of the extraction of knowledge, it is vital that these are of high quality to ensure robust predictions and sound analysis. Most countries have put in place strong policies for pension agencies to collect and provide relevant data to ensure transparency. As a result, the pension industry produces a high volume of data at high speed and in different varieties. Each of these (volume, speed, variety) separately makes preservation of data quality a challenging task, and considered together they make the preservation of quality a daunting endeavor. The ways that data are collected and stored are often not suitable to directly apply data science techniques on them. Data integration is another major issue, since different departments or agencies in the pension industry mostly generate raw data that are often not suitable for straightforward analysis. Different solutions have been proposed two tackle these problems, but many of these solutions require further refinement. For instance, semantic technologies such as knowledge graphs and ontologies can be used to structure data better and to provide semantics, while facilitating data sharing and integration. Other solutions include methods for designing data management systems with data ready for machine learning as output.

Privacy and security

Privacy and security issues should be considered when running data analysis in pension applications. First, the privacy of collected data is highly critical, because personal or critical business data can be involved. Second, the various data providing agencies often apply different data security policiess, so it is vital to consider how the various policies within a single data science system can be aligned.

Interpretability vs accuracy

Considering the characteristics of pension data, analytic algorithms should be able to handle large volumes. In other words, the pension industry requires algorithms that can analyze data that come from a variety of sources in real time. Many attempts have been made to address this issue. For example, deep learning algorithms, which are a form of neural networks, can be applied through online training to cope with real time application. Studies show that they can reach a high accuracy rate, provided that the data are sufficiently rich and that there is time for training. However, deep learning algorithms are sensitive to noisy data. Another issue is that neural network-based algorithms lack interpretation, that is, data scientists often cannot unambiguously interpret the model results.

6. Conclusion

Given that artificial intelligence and data science are relatively young fields, their application in the pension industry is still in an early phase. So there is much potential to adopt state-of-the-art methodologies for solving challenges in the pension industry. This article surveys the existing body of academic literature to summarize how data science is being currently leveraged to deal with issues related to pensions. The focus of most studies discussed here is on prediction tasks and chatbot development. There remains ample room for more unsupervised learning like clustering that uses the latest methodologies, NLP techniques to glean insight from text and social media data, and applications of reinforcement learning.

Our survey identifies three streams of applications of data science in the pensions context: a customer-focused approach, organizational process optimization, and personnel optimization. Customer-focused approaches are mostly applied to research on retirees or potential retirees, ranging from prediction of individual preferences to services provided. Organizational process optimization and personnel optimization focus on the pension providing agencies themselves and their workforce.

There is a vast amount of data that is hardly ever used. Data science can use these idle data and has the potential to, for example, detect societal groups with insufficient pension savings. Also, data science could be used to detect and aid customers who face diversification decisions regarding their savings.

There are still a number of challenges that need to be overcome for an effective application of data science in the pensions framework. Data collection and quality need to be standardized to enable straightforward integration into the data science workflow. Security and privacy protocols need to be properly laid out and addressed. Research into science techniques to enable the interpretability of blackbox data is a worthwhile area for future research.

References

Aguirre, J.A., Ladera, J.E., Castillo, B.D. and& Mayorga, S.A., 2019. Predictive analysis for calculating the valuation of the affiliated fund of a private pension system using machine learning techniques and tools. In the 2019 conference of Latin American and Caribbean Consortium of Engineering Institutions (LACCEI). *http://dx.doi.org/10.18687/LACCEI2019.1.1.343*.

Alavi, A.H., Gandomi, A.H. & Larry, D.J., 2016.The Progress of Machine Learning in Geosciences: Preface. Geoscience Frontiers 7(1): 21–31. *https://doi.org/10.1016/j.gsf.2015.10.006*.

Avramov, D., Cheng, S. & Metzker, L., 2019. Machine learning versus economic restrictions: Evidence from stock return predictability. Available at SSRN: *https://ssrn.com/abstract=3450322* or *http://dx.doi.org/10.2139/ssrn.3450322*.

Bacham, D. & Zhao, J., 2017. Machine Learning: Challenges, Lessons, and Opportunities in Credit Risk Modeling. Moody's analytics risk perspectives, managing disruption, Volume ix, July 2017. *https://www.moodysanalytics.com/risk-perspectives-magazine/managing-disruption/spotlight/machine-learning-challenges-lessons-and-opportunities-in-credit-risk-modeling*

Barber, D., 2012. Bayesian reasoning and machine learning. Cambridge University Press.

Bell, R., Koren, Y. & Volinsky, C., 2008. The BellKor 2008 Solution to the Netflix Prize.

Bengio, Y., Goodfellow, I.J. & Courville, A., 2016. Deep learning, MIT Press.

Bishop, C.M., 2006. Pattern recognition and machine learning. Springer.

Blekesaune, M. & Skirbekk, V., 2012."Can Personality Predict Retirement Behaviour? A Longitudinal Analysis Combining Survey and Register Data from Norway". European Journal of Ageing 9 (3): 199 – 206. *https://doi.org/10.1007/s10433-011-0212-3*.

Bolton, R.J. & Hand, D.J., 2001. Unsupervised Profiling Methods for Fraud Detection. Credit Scoring and Credit Control 7: 235–255. http://citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=EA8DC3FC4AD4C449E74F96 A7631DC2B1?doi=10.1.1.24.5743.

Bosworth, B. P. & Burtless, G., 2010. "Recessions, Wealth Destruction, and the Timing of Retirement". Working Paper 2010-22, Chestnut Hill, Center for Retirement Research at Boston College.

Caruana, R. & Niculescu-Mizil, A., 2006. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 161-168).

Coates, A. & Ng, A.Y., 2012. Learning feature representations with k-means. In Neural networks: Tricks of the trade (pp. 561-580). Springer, Berlin and Heidelberg.

Coile, C. & Levine, P.B., 2011. "The Market Crash and Mass Layoffs: How the Current Economic Crisis May Affect Retirement". The B.E. Journal of Economic Analysis and Policy 11(1): 22. doi:10.2202/1935-1682.2568.

Cong, J., Li, C., Liu, F. & Chu, D., 2016, October. EPF: An Elderly Personalization Features Based Collaborative Filtering Algorithm for Pension Service. In 2016 9th International Conference on Service Science (ICSS) (pp. 94-99).

Cong, L.W., Tang, K., Wang, J. & Zhang, Y., 2020. AlphaPortfolio for Investment and Economically Interpretable AI. Available at SSRN: *https://ssrn.com/abstract=3554486* or *http://dx.doi.org/10.2139/ssrn.3554486*.

Cortes, C. & Vapnik, V., 1995. Support-vector networks. Machine Learning, 20(3), pp.273-297.

Delamaire, L., Abdou, H. & Pointon, J. 2009."Credit Card Fraud and Detection Techniques: A Review". Banks and Bank Systems 4(2): pp. 57–68. *http://eprints.hud.ac.uk/id/eprint/19069*.

Deprez, P., Shevchenko, P.V. & Wuthrich, M.V., 2017."Machine Learning Techniques for Mortality Modeling". European Actuarial Journal 7: pp. 337–352. *https://doi.org/10.1007/s13385-017-0152-4*.

Dong, X., Li, C. & Chu, D., 2017, December. A Recommendation of Pension Service Based on Trusted Network. In 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC) (pp. 1251-1255).

Feldman, D.C. & Beehr, T.A., 2011. "A Three-Phase Model of Retirement Decision Making". American Psychologist 66(3):193-203. doi: 10.1037/a0022153.

Gough, O. & Sozou, P.D., 2005. Pensions and retirement savings: cluster analysis of consumer behaviour and attitudes. International Journal of Bank Marketing, 23(7), pp.558-570.

Hendriksen, A., 2017. Analytics on pension valuations. Research Paper Business Analytics.

Henkel, A.P., Bromuri, S., Iren, D. & Urovi, V., 2020. Half human, half machine – augmenting service employees with AI for interpersonal emotion regulation. Journal of Service Management.

Hosmer Jr, D.W., Lemeshow, S. & Sturdivant, R.X., 2013. Applied logistic regression (Vol. 398). John Wiley & Sons.

James, G., Witten, D., Hastie, T. & Tibshirani, R., 2013. *An introduction to statistical learning* (Vol. 112, p. 18). New York: Springer.

Jumutc, V., Langone, R. & Suykens, J.A., 2015, October. Regularized and sparse stochastic k-means for distributed large-scale clustering. In 2015 IEEE International Conference on Big Data (Big Data) (pp. 2535-2540). IEEE.

Juszczak, P., Adams, N.M., Hand, D.J., Whitrow, C. & Weston, D.J., 2008."The Peg and Bespoke Classifiers for Fraud Detection". Computational Statistics and Data Analysis 52(9):pp. 4521–4532. *https://doi.org/10.1016/j.csda.2008.03.014*.

Kaufmann, W., Starink, B. & Werker, B., 2018. Is de toekomst gearriveerd? Data science en individuele keuzemogelijkheden in pensioen. Netspar Industry Series. Design Paper 109.

Leung, W.S., Wong, W.K. & Wong, G., 2019. Social-Media Sentiment, Portfolio Complexity, and Stock Returns. SSRN (November 24, 2019). Available at SSRN: *https://ssrn.com/abstract=3492722* or *http://dx.doi.org/10.2139/ssrn.3492722*.

Li, C., Chu, D., Xu, X. & Bu, Y., 2019, July. A User Profile Based Pension Service Recommendation Algorithm. In 2019 IEEE World Congress on Services (SERVICES) (Vol. 2642, pp. 374-375).

Likas, A., Vlassis, N. & Verbeek, J.J., 2003. The global k-means clustering algorithm. Pattern recognition, 36(2), pp.451-461.

Lin T.C., 2001. "Letter to the Editor: Impact of Job Stress on Early Retirement Intention". International Journal of Stress Management 8(3): 243-247. *https://doi.org/10.1023/A:1011395227349*.

McAfee, A. & Brynjolfsson, E., 2012. Big Data: The Management Revolution. Harvard Business Review 90(10):60-6, 68, 128.

Murphy, K.P., 2012. Machine learning: a probabilistic perspective. MIT Press.

Otranto, E. & Trudda, A., 2008. Classifying Italian pension funds via GARCH distance. In Mathematical and Statistical Methods in Insurance and Finance (pp. 189-197). Springer, Milan.

Olaniyi, O., Ajibola, E., Ibiwoye, A. & Sogunro, A.B., 2012. Artificial Neural Network Model for Predicting Insolvency in Insurance Industry. International Journal of Management and Business Research 2(1): 59-68. *http://ijmbr.srbiau.ac.ir/article_63.html*.

Pozzolo, A.D., Caelen, O., Le Borgne, Y.A. & Waterschoot, S., 2014. Learned Lessons in Credit Card Fraud Detection from a Practitioner Perspective. Expert Systems with Applications 41(10): 4915–4928. *https://doi.org/10.1016/j.eswa.2014.02.026*.

Frisch, Ragnar. "Editor's Note." Econometrica, vol. 1, no. 1, 1933, pp. 1–4. JSTOR, www.jstor.org/stable/1912224. Accessed Feb. 4, 2021.

Rapach, D.E., Strauss, J.K., Tu, J. & Zhou, G., 2019. Industry return predictability: A machine learning approach. The Journal of Financial Data Science, 1(3), pp. 9–28.

Salazar, J.d.J.R. & Boado-Penas, M.D.C.B., 2019. Scoring and prediction of early retirement using machine learning techniques: application to private pensions plans. In Anales del Instituto de Actuarios Españoles (No. 25, pp. 119–145). Instituto de Actuarios Españoles.

Sasaki, T., Koizumi, H., Tajiri, T. & Kitano, H., 2018. A Study on the Use of Artificial Intelligence within Government Pension Investment Fund's Investment Management Practices (Summary Report). Tokyo, Japan: Government Pension Investment Fund. www.gpif.go.jp. Retrieved from *https://www.gpif.go.jp/en/investment/research_2017_1_en.pdf*.

Sayad, S., 2012. Logistic regression curve and Support vector machine, accessed August 2, 2020, <*https://www.saedsayad.com/logistic_regression.htm*>, <*http://www.saedsayad.com/support_vector_machine.htm*>.

Snow, D., 2019. Machine Learning in Asset Management. JFDS: *https://jfds.pm-research.com/content/2/1/10* , available at SSRN: *https://ssrn.com/abstract=3420952* or *http://dx.doi.org/10.2139/ssrn.3420952*

Talwar, A. & Kumar, Y., 2013. Machine Learning: An Artificial Intelligence Methodology. International Journal of Engineering and Computer Science 2(12): 3400–3404. *http://ijecs.in/index.php/ijecs/article/view/2261*.

Werpachowska, A., 2018. Forecasting the impact of state pension reforms in post-Brexit England and Wales using microsimulation and deep learning. arXiv preprint arXiv:1802.09427.

OVERZICHT UITGAVEN IN DE DESIGN PAPER SERIE

1 Naar een nieuw pensioencontract (2011)
Lans Bovenberg en Casper van Ewijk
2 Langlevenrisico in collectieve pensioencontracten (2011)
Anja De Waegenaere, Alexander Paulis en Job Stigter
3 Bouwstenen voor nieuwe pensioencontracten en uitdagingen voor het toezicht daarop (2011)
Theo Nijman en Lans Bovenberg
4 European supervision of pension funds: purpose, scope and design (2011)
Niels Kortleve, Wilfried Mulder and Antoon Pelsser
5 Regulating pensions: Why the European Union matters (2011)
Ton van den Brink, Hans van Meerten and Sybe de Vries
6 The design of European supervision of pension funds (2012)
Dirk Broeders, Niels Kortleve, Antoon Pelsser and Jan-Willem Wijckmans
7 Hoe gevoelig is de uittredeleeftijd voor veranderingen in het pensioenstelsel? (2012)
Didier Fouarge, Andries de Grip en Raymond Montizaan
8 De inkomensverdeling en levensverwachting van ouderen (2012)
Marike Knoef, Rob Alessie en Adriaan Kalwij
9 Marktconsistente waardering van zachte pensioenrechten (2012)
Theo Nijman en Bas Werker
10 De RAM in het nieuwe pensioenakkoord (2012)
Frank de Jong en Peter Schotman
11 The longevity risk of the Dutch Actuarial Association's projection model (2012)
Frederik Peters, Wilma Nusselder and Johan Mackenbach
12 Het koppelen van pensioenleeftijd en pensioenaanspraken aan de levensverwachting (2012)
Anja De Waegenaere, Bertrand Melenberg en Tim Boonen
13 Impliciete en expliciete leeftijdsdifferentiatie in pensioencontracten (2013)
Roel Mehlkopf, Jan Bonenkamp, Casper van Ewijk, Harry ter Rele en Ed Westerhout
14 Hoofdlijnen Pensioenakkoord, juridisch begrepen (2013)
Mark Heemskerk, Bas de Jong en René Maatman
15 Different people, different choices: The influence of visual stimuli in communication on pension choice (2013)
Elisabeth Brüggen, Ingrid Rohde and Mijke van den Broeke
16 Herverdeling door pensioenregelingen (2013)
Jan Bonenkamp, Wilma Nusselder, Johan Mackenbach, Frederik Peters en Harry ter Rele
17 Guarantees and habit formation in pension schemes: A critical analysis of the floor-leverage rule (2013)
Frank de Jong and Yang Zhou
18 The holistic balance sheet as a building block in pension fund supervision (2013)
Erwin Fransen, Niels Kortleve, Hans Schumacher, Hans Staring and Jan-Willem Wijckmans
19 Collective pension schemes and individual choice (2013)
Jules van Binsbergen, Dirk Broeders, Myrthe de Jong and Ralph Koijen
20 Building a distribution builder: Design considerations for financial investment and pension decisions (2013)
Bas Donkers, Carlos Lourenço, Daniel Goldstein and Benedict Dellaert

21 Escalerende garantietoezeggingen: een alternatief voor het StAr RAM-contract (2013)
Servaas van Bilsen, Roger Laeven en Theo Nijman
22 A reporting standard for defined contribution pension plans (2013)
Kees de Vaan, Daniele Fano, Herialt Mens and Giovanna Nicodano
23 Op naar actieve pensioenconsumenten: Inhoudelijke kenmerken en randvoorwaarden van effectieve pensioencommunicatie (2013)
Niels Kortleve, Guido Verbaal en Charlotte Kuiper
24 Naar een nieuw deelnemergericht UPO (2013)
Charlotte Kuiper, Arthur van Soest en Cees Dert
25 Measuring retirement savings adequacy; developing a multi-pillar approach in the Netherlands (2013)
Marike Knoef, Jim Been, Rob Alessie, Koen Caminada, Kees Goudswaard, and Adriaan Kalwij
26 Illiquiditeit voor pensioenfondsen en verzekeraars: Rendement versus risico (2014)
Joost Driessen
27 De doorsneesystematiek in aanvullende pensioenregelingen: effecten, alternatieven en transitiepaden (2014)
Jan Bonenkamp, Ryanne Cox en Marcel Lever
28 EIOPA: bevoegdheden en rechtsbescherming (2014)
Ivor Witte
29 Een institutionele beleggersblik op de Nederlandse woningmarkt (2013)
Dirk Brounen en Ronald Mahieu
30 Verzekeraar en het reële pensioencontract (2014)
Jolanda van den Brink, Erik Lutjens en Ivor Witte
31 Pensioen, consumptiebehoeften en ouderenzorg (2014)
Marike Knoef, Arjen Hussem, Arjan Soede en Jochem de Bresser
32 Habit formation: implications for pension plans (2014)
Frank de Jong and Yang Zhou
33 Het Algemeen pensioenfonds en de taakafbakening (2014)
Ivor Witte
34 Intergenerational Risk Trading (2014)
Jiajia Cui and Eduard Ponds
35 Beëindiging van de doorsneesystematiek: juridisch navigeren naar alternatieven (2015)
Dick Boeijen, Mark Heemskerk en René Maatman
36 Purchasing an annuity: now or later? The role of interest rates (2015)
Thijs Markwat, Roderick Molenaar and Juan Carlos Rodriguez
37 Entrepreneurs without wealth? An overview of their portfolio using different data sources for the Netherlands (2015)
Mauro Mastrogiacomo, Yue Li and Rik Dillingh
38 The psychology and economics of reverse mortgage attitudes. Evidence from the Netherlands (2015)
Rik Dillingh, Henriëtte Prast, Mariacristina Rossi and Cesira Urzì Brancati
39 Keuzevrijheid in de uittreedleeftijd (2015)
Arthur van Soest
40 Afschaffing doorsneesystematiek: verkenning van varianten (2015)
Jan Bonenkamp en Marcel Lever
41 Nederlandse pensioenopbouw in internationaal perspectief (2015)
Marike Knoef, Kees Goudswaard, Jim Been en Koen Caminada
42 Intergenerationele risicodeling in collectieve en individuele pensioencontracten (2015)
Jan Bonenkamp, Peter Broer en Ed Westerhout
43 Inflation Experiences of Retirees (2015)
Adriaan Kalwij, Rob Alessie, Jonathan Gardner and Ashik Anwar Ali
44 Financial fairness and conditional indexation (2015)
Torsten Kleinow and Hans Schumacher
45 Lessons from the Swedish occupational pension system (2015)
Lans Bovenberg, Ryanne Cox and Stefan Lundbergh

46 Heldere en harde pensioenrechten onder een PPR (2016)
Mark Heemskerk, René Maatman en Bas Werker
47 Segmentation of pension plan participants: Identifying dimensions of heterogeneity (2016)
Wiebke Eberhardt, Elisabeth Brüggen, Thomas Post and Chantal Hoet
48 How do people spend their time before and after retirement? (2016)
Johannes Binswanger
49 Naar een nieuwe aanpak voor risicoprofielmeting voor deelnemers in pensioenregelingen (2016)
Benedict Dellaert, Bas Donkers, Marc Turlings, Tom Steenkamp en Ed Vermeulen
50 Individueel defined contribution in de uitkeringsfase (2016)
Tom Steenkamp
51 Wat vinden en verwachten Nederlanders van het pensioen? (2016)
Arthur van Soest
52 Do life expectancy projections need to account for the impact of smoking? (2016)
Frederik Peters, Johan Mackenbach en Wilma Nusselder
53 Effecten van gelaagdheid in pensioendocumenten: een gebruikersstudie (2016)
Louise Nell, Leo Lentz en Henk Pander Maat
54 Term Structures with Converging Forward Rates (2016)
Michel Vellekoop and Jan de Kort
55 Participation and choice in funded pension plans (2016)
Manuel García-Huitrón and Eduard Ponds
56 Interest rate models for pension and insurance regulation (2016)
Dirk Broeders, Frank de Jong and Peter Schotman
57 An evaluation of the nFTK (2016)
Lei Shu, Bertrand Melenberg and Hans Schumacher
58 Pensioenen en inkomensongelijkheid onder ouderen in Europa (2016)
Koen Caminada, Kees Goudswaard, Jim Been en Marike Knoef
59 Towards a practical and scientifically sound tool for measuring time and risk preferences in pension savings decisions (2016)
Jan Potters, Arno Riedl and Paul Smeets
60 Save more or retire later? Retirement planning heterogeneity and perceptions of savings adequacy and income constraints (2016)
Ron van Schie, Benedict Dellaert and Bas Donkers
61 Uitstroom van oudere werknemers bij overheid en onderwijs. Selectie uit de poort (2016)
Frank Cörvers en Janneke Wilschut
62 Pension risk preferences. A personalized elicitation method and its impact on asset allocation (2016)
Gosse Alserda, Benedict Dellaert, Laurens Swinkels and Fieke van der Lecq
63 Market-consistent valuation of pension liabilities (2016)
Antoon Pelsser, Ahmad Salahnejhad and Ramon van den Akker
64 Will we repay our debts before retirement? Or did we already, but nobody noticed? (2016)
Mauro Mastrogiacomo
65 Effectieve ondersteuning van zelfmanagement voor de consument (2016)
Peter Lapperre, Alwin Oerlemans en Benedict Dellaert
66 Risk sharing rules for longevity risk: impact and wealth transfers (2017)
Anja De Waegenaere, Bertrand Melenberg and Thijs Markwat
67 Heterogeniteit in doorsneeproblematiek. Hoe pakt de transitie naar degressieve opbouw uit voor verschillende pensioenfondsen? (2017)
Loes Frehen, Wouter van Wel, Casper van Ewijk, Johan Bonekamp, Joost van Valkengoed en Dick Boeijen
68 De toereikendheid van pensioenopbouw na de crisis en pensioenhervormingen (2017)
Marike Knoef, Jim Been, Koen Caminada, Kees Goudswaard en Jason Rhuggenaath

69 De combinatie van betaald en onbetaald werk in de jaren voor pensioen (2017)
Marleen Damman en Hanna van Solinge

70 Default life-cycles for retirement savings (2017)
Anna Grebenchtchikova, Roderick Molenaar, Peter Schotman en Bas Werker

71 Welke keuzemogelijkheden zijn wenselijk vanuit het perspectief van de deelnemer? (2017)
Casper van Ewijk, Roel Mehlkopf, Sara van den Bleeken en Chantal Hoet

72 Activating pension plan participants: investment and assurance frames (2017)
Wiebke Eberhardt, Elisabeth Brüggen, Thomas Post en Chantal Hoet

73 Zerotopia – bounded and unbounded pension adventures (2017)
Samuel Sender

74 Keuzemogelijkheden en maatwerk binnen pensioenregelingen (2017)
Saskia Bakels, Agnes Joseph, Niels Kortleve en Theo Nijman

75 Polderen over het pensioenstelsel. Het debat tussen de sociale partners en de overheid over de oudedagvoorzieningen in Nederland, 1945-2000 (2017)
Paul Brusse

76 Van uitkeringsovereenkomst naar PPR (2017)
Mark Heemskerk, Kees Kamminga, René Maatman en Bas Werker

77 Pensioenresultaat bij degressieve opbouw en progressieve premie (2017)
Marcel Lever en Sander Muns

78 Bestedingsbehoeften bij een afnemende gezondheid na pensionering (2017)
Lieke Kools en Marike Knoef

79 Model Risk in the Pricing of Reverse Mortgage Products (2017)
Anja De Waegenaere, Bertrand Melenberg, Hans Schumacher, Lei Shu and Lieke Werner

80 Expected Shortfall voor toezicht op verzekeraars: is het relevant? (2017)
Tim Boonen

81 The Effect of the Assumed Interest Rate and Smoothing on Variable Annuities (2017)
Anne G. Balter and Bas J.M. Werker

82 Consumer acceptance of online pension investment advice (2017)
Benedict Dellaert, Bas Donkers and Carlos Lourenço

83 Individualized life-cycle investing (2017)
Gréta Oleár, Frank de Jong and Ingmar Minderhoud

84 The value and risk of intergenerational risk sharing (2017)
Bas Werker

85 Pensioenwensen voor en na de crisis (2017)
Jochem de Bresser, Marike Knoef en Lieke Kools

86 Welke vaste dalingen en welk beleggingsbeleid passen bij gewenste uitkeringsprofielen in verbeterde premieregelingen? (2017)
Johan Bonekamp, Lans Bovenberg, Theo Nijman en Bas Werker

87 Inkomens- en vermogensafhankelijke eigen bijdragen in de langdurige ouderenzorg: een levensloopperspectief (2017)
Arjen Hussem, Harry ter Rele en Bram Wouterse

88 Creating good choice environments – Insights from research and industry practice (2017)
Elisabeth Brüggen, Thomas Post and Kimberley van der Heijden

89 Two decades of working beyond age 65 in the Netherlands. Health trends and changes in socio-economic and work factors to determine the feasibility of extending working lives beyond age 65 (2017)
Dorly Deeg, Maaike van der Noordt and Suzan van der Pas

90 Cardiovascular disease in older workers. How can workforce participation be maintained in light of changes over time in determinants of cardiovascular disease? (2017)
Dorly Deeg, E. Burgers and Maaike van der Noordt

91 Zicht op zzp-pensioen (2017)
Wim Zwinkels, Marike Knoef, Jim Been, Koen Caminada en Kees Goudswaard

92 Return, risk, and the preferred mix of PAYG and funded pensions (2017)
Marcel Lever, Thomas Michielsen and Sander Muns

93 Life events and participant engagement in pension plans (2017)
Matthew Blakstad, Elisabeth Brüggen and Thomas Post
94 Parttime pensioneren en de arbeidsparticipatie (2017)
Raymond Montizaan
95 Keuzevrijheid in pensioen: ons brein wil niet kiezen, maar wel gekozen hebben (2018)
Walter Limpens en Joyce Vonken
96 Employability after age 65? Trends over 23 years in life expectancy in good and in poor physical and cognitive health of 65-74-year-olds in the Netherlands (2018)
Dorly Deeg, Maaike van der Noordt, Emiel Hoogendijk, Hannie Comijs and Martijn Huisman
97 Loslaten van de verplichte pensioenleeftijd en het organisatieklimaat rondom langer doorwerken (2018)
Jaap Oude Mulders, Kène Henkens en Harry van Dalen
98 Overgangseffecten bij introductie degressieve opbouw (2018)
Bas Werker
99 You're invited – RSVP! The role of tailoring in incentivising people to delve into their pension situation (2018)
Milena Dinkova, Sanne Elling, Adriaan Kalwij en Leo Lentz
100 Geleidelijke uittreding en de rol van deeltijdpensioen (2018)
Jonneke Bolhaar en Daniël van Vuuren
101 Naar een model voor pensioencommunicatie (2018)
Leo Lentz, Louise Nell en Henk Pander Maat
102 Tien jaar UPO. Een terugblik en vooruitblik op inhoud, doelen en effectiviteit (2018)
Sanne Elling en Leo Lentz
103 Health and household expenditures (2018)
Raun van Ooijen, Jochem de Bresser en Marike Knoef
104 Keuzevrijheid in de uitkeringsfase: internationale ervaringen (2018)
Marcel Lever, Eduard Ponds, Rik Dillingh en Ralph Stevens
105 The move towards riskier pension products in the world's best pension systems (2018)
Anne G. Balter, Malene Kallestrup-Lamb and Jesper Rangvid
106 Life Cycle Option Value: The value of consumer flexibility in planning for retirement (2018)
Sonja Wendel, Benedict Dellaert and Bas Donkers
107 Naar een duidelijk eigendomsbegrip (2018)
Jop Tangelder
108 Effect van stijging AOW-leeftijd op arbeidsongeschiktheid (2018)
Rik Dillingh, Jonneke Bolhaar, Marcel Lever, Harry ter Rele, Lisette Swart en Koen van der Ven
109 Is de toekomst gearriveerd? Data science en individuele keuzemogelijkheden in pensioen (2018)
Wesley Kaufmann, Bastiaan Starink en Bas Werker
110 De woontevredenheid van ouderen in Nederland (2018)
Jan Rouwendal
111 Towards better prediction of individual longevity (2018)
Dorly Deeg, Jan Kardaun, Maaike van der Noordt, Emiel Hoogendijk en Natasja van Schoor
112 Framing in pensioenkeuzes. Het effect van framing in de keuze voor beleggingsprofiel in DC-plannen naar aanleiding van de Wet verbeterde premieregeling (2018)
Marijke van Putten, Rogier Potter van Loon, Marc Turlings en Eric van Dijk
113 Working life expectancy in good and poor self-perceived health among Dutch workers aged 55–65 years with a chronic disease over the period 1992–2016 (2019)
Astrid de Wind, Maaike van der Noordt, Dorly Deeg and Cécile Boot
114 Working conditions in post-retirement jobs: A European comparison (2019)
Ellen Dingemans and Kène Henkens

115 Is additional indebtedness the way to increase mortgage-default insurance coverage? (2019)
Yeorim Kim, Mauro Mastrogiacomo, Stefan Hochguertel and Hans Bloemen

116 Appreciated but complicated pension Choices? Insights from the Swedish Premium Pension System (2019)
Monika Böhnke, Elisabeth Brüggen and Thomas Post

117 Towards integrated personal financial planning. Information barriers and design propositions (2019)
Nitesh Bharosa and Marijn Janssen

118 The effect of tailoring pension information on navigation behavior (2019)
Milena Dinkova, Sanne Elling, Adriaan Kalwij and Leo Lentz

119 Opleiding, levensverwachting en pensioenleeftijd: een vergelijking van Nederland met andere Europese landen (2019)
Johan Mackenbach, José Rubio Valverde en Wilma Nusselder

120 Giving with a warm hand: Evidence on estate planning and bequests (2019)
Eduard Suari-Andreu, Raun van Ooijen, Rob J.M. Alessie and Viola Angelini

121 Investeren in menselijk kapitaal: een gecombineerd werknemers- en werkgeversperspectief (2019)
Raymond Montizaan, Merlin Nieste en Davey Poulissen

122 The rise in life expectancy – corresponding rise in subjective life expectancy? Changes over the period 1999-2016 (2019)
Dorly Deeg, Maaike van der Noordt, Noëlle Sant, Henrike Galenkamp, Fanny Janssen and Martijn Huisman

123 Pensioenaanvullingen uit het eigen woningbezit (2019)
Dirk Brounen, Niels Kortleve en Eduard Ponds

124 Personal and work-related predictors of early exit from paid work among older workers with health limitations (2019)
Nils Plomp, Sascha de Breij and Dorly Deeg

125 Het delen van langlevenrisico (2019)
Anja De Waegenaere, Agnes Joseph, Pascal Janssen en Michel Vellekoop

126 Maatwerk in pensioencommunicatie (2019)
S.K. Elling en L.R. Lentz

127 Dutch Employers' Responses to an Aging Workforce: Evidence from Surveys, 2009-2017 (2019)
Jaap Oude Mulders, Kène Henkens and Hendrik P. van Dalen

128 Preferences for solidarity and attitudes towards the Dutch pension system – Evidence from a representative sample (2019)
Arno Riedl, Hans Schmeets and Peter Werner

129 Deeltijdpensioen geen wondermiddel voor langer doorwerken (2019)
Henk-Wim de Boer, Tunga Kantarcı, Daniel van Vuuren en Ed Westerhout

130 Spaarmotieven en consumptiegedrag (2019)
Johan Bonekamp en Arthur van Soest

131 Substitute services: a barrier to controlling long-term care expenditures (2019)
Mark Kattenberg and Pieter Bakx

132 Voorstel keuzearchitectuur pensioensparen voor zelfstandigen (2019)
Jona Linde

133 The impact of the virtual integration of assets on pension risk preferences of individuals (2019)
Sesil Lim, Bas Donkers en Benedict Dellaert

134 Reforming the statutory retirement age: Policy preferences of employers (2019)
Hendrik P. van Dalen, Kène Henkens and Jaap Oude Mulders

135 Compensatie bij afschaffing doorsnee-systematiek (2019)
Dick Boeijen, Chantal de Groot, Mark Heemskerk, Niels Kortleve en René Maatman

136 Debt affordability after retirement, interest rate shocks and voluntary repayments (2019)
Mauro Mastrogiacomo

137 Using social norms to activate pension plan members: insights from practice (2019)
Joyce Augustus-Vonken, Pieter Verhallen, Lisa Brüggen and Thomas Post

138 Alternatieven voor de huidige verplichtstelling van bedrijfstakpensioenfondsen (2020)
Erik Lutjens en Fieke van der Lecq

139 Eigen bijdrage aan ouderenzorg (2020)
Pieter Bakx, Judith Bom, Marianne Tenand en Bram Wouterse

140 Inrichting fiscaal kader bij afschaffing doorsneesystematiek (2020)
Bastiaan Starink en Michael Visser

141 Hervorming langdurige zorg: trends in het gebruik van verpleging en verzorging (2020)
Pieter Bakx, Pilar Garcia-Gomez, Sara Rellstab, Erik Schut en Eddy van Doorslaer

142 Genetic health risks, insurance, and retirement (2020)
Richard Karlsson Linnér and Philipp D. Koellinger

143 Publieke middelen voor particuliere ouderenzorg (2020)
Arjen Hussem, Marianne Tenand en Pieter Bakx

144 Emotions and technology in pension service interactions: Taking stock and moving forward (2020)
Wiebke Eberhardt, Alexander Henkel en Chantal Hoet

145 Opleidingsverschillen in levensverwachting: de bijdrage van acht risicofactoren (2020)
Wilma J. Nusselder, José Rubio Valverde en Johan P. Mackenbach

146 Shades of Labor: Motives of Older Adults to Participate in Productive Activities (2020)
Sonja Wendel and Benedict Dellaert

147 Raising pension awareness through letters and social media: Evidence from a randomized and a quasi-experiment (2020)
Marike Knoef, Jim Been and Marijke van Putten

148 Infographics and Financial Decisions (2020)
Ruben Cox and Peter de Goeij

149 To what extent can partial retirement ensure retirement income adequacy? (2020)
Tunga Kantarcı and Jochem Zweerink

150 De steun voor een 'zwareberoepenregeling' ontleed (2020)
Harry van Dalen, Kène Henkens en Jaap Oude Mulders

151 Verbeteren van de inzetbaarheid van oudere werknemers tot aan pensioen: literatuuroverzicht, inzichten uit de praktijk en de rol van pensioenuitvoerders (2020)
Peter Lapperre, Henk Heek, Pascal Corten, Ad van Zonneveld, Robert Boulogne, Marieke Koeman en Benedict Dellaert

152 Betere risicospreiding van eigen bijdragen in de verpleeghuiszorg (2020)
Bram Wouterse, Arjen Hussem en Rob Aalbers

153 Doorbeleggen met garanties? (2020)
Roderick Molenaar, Peter Schotman, Peter Dekkers en Mark Irwin

154 Differences in retirement preferences between the self-employed and employees: Do job characteristics play an explanatory role? (2020)
Marleen Damman, Dieuwke Zwier en Swenne G. van den Heuvel

155 Do financial incentives stimulate partially disabled persons to return to work? (2020)
Tunga Kantarcı and Jan-Maarten van Sonsbeek

156 Wijzigen van de bedrijfstakpensioenregeling: tussen pensioenfondsbestuur en sociale partners (2020)
J.R.C. Tangelder

157 Keuzes tijdens de pensioenopbouw: de effecten van nudging met volgorde en standaardopties (2020)
Wilte Zijlstra, Jochem de Bresser en Marike Knoef

158 Keuzes rondom pensioen: implicaties op uitkeringssnelheid voor een heterogeen deelnemersbestand (2020)
Servaas van Bilsen, Johan Bonekamp, en Eduard Ponds

159 Met big data inspelen op woonwensen en woongedrag van ouderen: praktische inzichten voor ontwerp en beleid (2020)
Ioulia V. Ossokina en Theo A. Arentze

160 Economic consequences of widowhood: Evidence from a survivor's benefits reform in the Netherlands (2020)
Jeroen van der Vaart, Rob Alessie and Raun van Ooijen

161 How will disabled workers respond to a higher retirement age? (2020)
Tunga Kantarcı, Jim Been and Arthur van Soest

162 Deeltijdpensioen: belangstelling en belemmeringen op de werkvloer (2020)
Hanna van Solinge, Harry van Dalen en Kène Henkens

163 Investing for Retirement with an Explicit Benchmark (2020)
Anne Balter, Lennard Beijering, Pascal Janssen, Frank de Jong, Agnes Joseph, Thijs Kamma and Antoon Pelsser

164 Vergrijzing en verzuim: impact op de verzekeringsvoorkeuren van werkgevers (2020)
Remco Mallee en Raymond Montizaan

165 Arbeidsmarkteffecten van de pensioen-premiesystematiek (2020)
Marike Knoef, Sander Muns en Arthur van Soest

166 Risk Sharing within Pension Schemes (2020)
Anne Balter, Frank de Jong en Antoon Pelsser

167 Supporting pension participants: Three lessons learned from the medical domain for better pension decisions (2021)
Jelle Strikwerda, Bregje Holleman and Hans Hoeken

168 Variable annuities with financial risk and longevity risk in the decumulation phase of Dutch DC products (2021)
Bart Dees, Frank de Jong and Theo Nijman

169 Verloren levensjaren als gevolg van sterfte aan Covid-19 (2021)
Bram Wouterse, Frederique Ram en Pieter van Baal

170 Which work conditions can encourage older workers to work overtime? (2021)
Raymond Montizaan and Annemarie Kuenn-Nelen

171 Herverdeling van individueel pensioen-vermogen naar partnerpensioen: een stated preference-analyse (2021)
Raymond Montizaan

172 Risicogedrag na een ramp; implicaties voor pensioenen (2021)
Martijn de Vries

173 The Impact of Climate Change on Optimal Asset Allocation for Long-Term Investors (2021)
Mathijs Cosemans, Xander Hut and Mathijs van Dijk

174 Beleggingsbeleid bij onzekerheid over risicobereidheid en budget (2021)
Agnes Joseph, Antoon Pelsser en Lieke Werner

175 On the Resilience of ESG Stocks during COVID-19: Global Evidence (2021)
Gianfranco Gianfrate, Tim Kievid & Mathijs van Dijk

176 De solidariteitsreserve juridisch ontrafeld (2021)
Erik Lutjens en Herman Kappelle

177 Hoe vertrouwen in politiek en maatschappij doorwerkt in vertrouwen in pensioen-instituties (2021)
Harry van Dalen en Kène Henkens

178 Gelijke rechten, maar geen gelijke pensioenen: de gender gap in Nederlandse tweedepijlerpensioenen (2021)
Suzanne Kali, Jim Been, Marike Knoef en Albert van Marwijk Kooy

179 Completing Dutch pension reform (2021)
Ed Westerhout, Eduard Ponds and Peter Zwaneveld

180 When and why do employers hire and rehire employees beyond normal retirement age? (2021)
Orlaith C. Tunney and Jaap Oude Mulders

181 Family and government insurance: Wage, earnings, and income risks in the Netherlands and the U.S. (2021)
Mariacristina De Nardi, Giulio Fella, Marike Knoef, Gonzalo Paz-Pardo and Raun van Ooijen

Network for Studies on Pensions, Aging and Retirement

This is a publication of:
Netspar
Phone +31 13 466 2109
E-mail info@netspar.nl
www.netspar.nl

September 2021