

## EIT Digital – Industrial PhD position proposal

### PhD thesis information

PhD Thesis – Title	Customer profiling-based personalization in financial systems
PhD Thesis – Short summary	Modern financial institutions collect vast amount of historical data about the activities of their customers, e.g. number and type of interactions with the financial services provided, financial transaction details. The main task of the PhD project is to analyse those rich data sources, create algorithms and solutions which effectively support decision making processes by exploring clients’ preferences and forecasting their future behaviour. These goals will be achieved by creating and maintaining up-to-date customer profiles and developing algorithms and solutions which propose the optimal set of services (e.g. targeted marketing, next-product-to-buy) to customers based on their profiles.
Rationale/challenge – <i>describe the problem and why it is relevant</i>	The technological revolution of the past decades caused a high-speed accumulation of business related data in the retail banking sector; this trend is expected to accelerate during the coming years due to the spreading of smart appliances and online banking. Most importantly, clients’ behaviour is now recorded in more detail than ever before, with thousands of features about their demographic background, savings, investment portfolio, payment, browsing and purchasing history, and the financial products they own. The massive volume of client data – combined with external data sources like satellite images, maps, macroeconomic indices etc. – requires new analytic approaches to create useful insights for decision makers and develop/implement automated tools to collect, clean, validate, analyse data and generate real time recommendations. New methods are constantly developed in the data science community, but they are mostly focused on engineering problems like image and speech recognition. The challenge is how to use these techniques for retail banking purposes, i.e. which methods and approaches have a value added in the financial sector. For successful implementation not only methodologies matter, but infrastructural environments – e.g. how to implement and use ‘Big Data’ clusters – have a significant impact on usability as well. Another important criterion is the implementation of industry-grade solutions, which are capable to adapt to modified operational environments and to recalibrate model parameters when new behaviour patterns emerge in the constantly changing financial environment.
Innovation – <i>describe what is the intended solution and the advance w.r.t. the state-of-the-art</i>	Large banks – like OTP Bank Hungary – have millions of customers and thousands of potential features for each of them. Additionally, many of these features are time-dependent which makes it unavoidable to leverage the latest data science and data engineering solutions. In this doctoral work the state-of-the-art will be advanced in the customer profiling domain, allowing larger banks to create constantly evolving customer micro-profiles, which will allow them to provide personalized services, thereby optimizing their interactions with customers, offering the best next product and avoiding churn.

	<p>The research questions of the PhD will be formulated in line with the bank's need to identify clients' future behaviour, which is a true pain point, whose solution has the potential to revolutionize the interaction between the bank and its customers. A draft list of novel research questions which will be analysed during the PhD project follows:</p> <ul style="list-style-type: none"> <li>• Churn prediction/survival analysis. (Estimating the probability that a customer will leave the bank, won't extend the contract after expiration date etc. Implementing an early warning system that is constantly searching and automatically signalling if a client has high 'churn risk'.)</li> <li>• Recommendation systems. (Developing algorithms that are apt at real time identification of which product/service should be recommended for the client based on her past behaviour. The optimal choice must be generated in terms of customer lifetime value, market share, utility etc.)</li> <li>• Data visualization. (Visualization of the results in a user-friendly way via charts, interactive dashboards etc.)</li> </ul>
<p>Research focus/topics – describe <u>how</u> you are going to solve the problem</p>	<p>The most important research topics will be the following:</p> <ol style="list-style-type: none"> <li>1) Collect information about the financial system, specify the most important pain points in behaviour prediction and formulate an initial set of hypotheses.</li> <li>2) Identifying business/analytical needs that will evolve in the near future as a consequence of technological progress, new data sources and changing business environment. Explore and implement current state-of-the-art solutions and on-going researches in the topic.</li> <li>3) Create the first prototypes of models and automated data analysis and prediction solutions. Test them on real datasets provided by OTP.</li> <li>4) Improve the prototypes and deploy them in OTP production environment, or in an environment which is populated by real time financial data about customers.</li> <li>5) Evaluate the proper operation of the implemented solutions and make further developments based on empirical evidence, e.g. accuracy, customer satisfaction, financial impact.</li> </ol>
<p>Deadlines/milestones (Gantt chart)</p>	<p>Review industrial best practices, current solutions and most recent research developments. Understand the business environment and business needs via analysing the business processes and their underlying datasets. Specify the most important pain points and create a list of hypotheses. The hypotheses must cover the following topics: churn prediction, client segmentation and optimal recommendation in multiple choice systems.</p> <p>Implement state-of-the-art baseline solutions developed by leading experts for clustering, classification and prediction problems. Finalize the hypotheses with OTP.</p>

	Automate data collection, cleaning and structuring. Completion of the milestone requires the processing of external and internal (OTP) data sources as well. Explorative data analysis, client profiling. Receive feedback from OTP eBIZ, where the customer (micro-)profiles will be utilized as part of their targeted marketing activities.
	Develop and test the first prototypes of the predictive models. Performance must be measured not only by standard statistical metrics but by expected business impact.
	Validate and verify the models in a production environment. Optimize the prototypes. Ensure that they are suitable to be industrialized in the OTP environment. Measure the success rate of the solution via at least two real-life experiments, e.g. acquisition of new customers via targeted marketing, optimizing the next-product-to-buy strategy, reducing customer churn.
	Summarize the results and complete the PhD thesis.
Expected outcome – <i>describe the expected results of the PhD</i>	<p>The results of the PhD will be a toolbox with implemented algorithms able to automatically collect, structure and clean customer data, give model-based recommendations and predictions. It will be implemented in a financial system’s production environment (i.e. OTP Bank) as a business supporting tool.</p> <p>The measurable outputs shall include:</p> <ul style="list-style-type: none"> <li>• Documentation of existing industrial solutions in the field of research.</li> <li>• Implementation of existing solutions utilized as baselines.</li> <li>• Implemented behaviour prediction toolbox.</li> <li>• A detailed analysis of the candidate’s contribution and the impact on the bank’s operational and business processes.</li> <li>• Published papers in journals and conferences.</li> </ul>

Relevance for the Action Line (section to be filled out by the Action Line Leader)

Action Line	Digital Finance
Alignment with Action Line – <i>statement from the Action Line Leader indicating the relevance for the AL from his perspective</i>	...
Relevant IA – <i>List any relevant Innovation Activity (if applicable)</i>	...

Partnership/financial information

Action Line Leader	Antonio Garcia Hortal
Industrial partner	OTP Bank

Industry advisor – <i>name and short bio</i>	<p>Illés Gozlán</p> <p>Illés Gozlán is the Head of Data Science and Customer Value Optimization at OTP Bank. He received an MSc in Economics (major in Information Technology and Decision Theory) at Corvinus University of Budapest in 2000 and a BSc in Software Engineering at Széchenyi István University in 2004. He worked 5 years as a data mining consultant at Data Explorer Inc. between 2000 and 2004 where his research area mainly covered the Telecommunication, Pharmaceutical and Banking industries. He was also a senior Data Scientist and Project Manager in several projects at T-Mobile USA – Seattle. He had joined OTP Bank in 2011 and since then he developed a versatile Data Science team of 18 experts from various fields. Besides data scientists his team consists of database and data management experts, economists, project managers, developers and business analysts as well. The holistic data and analytical capabilities of the team aim that all kind of data driven problems can be handled effectively inside OTP Group which is one of the largest financial service providers in Central and Eastern Europe with more than 17 million clients.</p>
Academic/research partner	ELTE
Academic/research supervisor – <i>name and short bio</i>	<p>Imre Lendák, PhD</p> <p>Imre Lendák is an assistant professor at the Eötvös Loránd University (ELTE). He obtained his PhD from the University of Novi Sad (Serbia) in 2011 for developing a data analysis algorithm for identifying repetitive topologies in large network models of electric power distribution systems. His current research interests include applied security data science in critical infrastructures and graph visualization. He coordinates one Erasmus+ Capacity Building in Higher Education (CBHE) project with the goal to develop different MSc and specialization programs in information security at four higher education institutions in Serbia. He is an IEEE and ACM member.</p>
HEI granting the title	ELTE
DTC location	Budapest DTC
Geographical mobility plan	
No. of PhD positions	1
PhD duration	4
Co-funding percentages:	20
- Industry	30
- Academia	50
- EIT Digital	