EIT Digital – Industrial PhD position proposal

PhD thesis information

PhD Thesis – Title	Optimization of internal processes in financial systems
PhD Thesis – Short summary	The vast amount of structured and unstructured data that is generated by the internal processes of modern financial institutions requires efficient solutions for automated data preparation (collecting, cleaning), modelling, analysis and predictive modelling. The main task of this PhD is to develop process- and data mining algorithms, which will enable financial systems to significantly optimize their internal processes on one hand, while maintaining the current information security posture and protection of personally identifiable information (PII). This requires an in-depth understanding of the way companies in this field operate. The industrial partner will guide the doctoral student through business process identification, cost measurement and optimization.
Rationale/challenge – *describe the problem and why it is relevant*	Modern financial institutions are faced with the challenges of collecting, cleaning, storing and analysing large volumes of information generated by their business processes. Apart from recording information about their interactions with their customers, modern banks are also equipped with the necessary tools to record rich sets of information about their internal business processes in the form of documentation, application logs, transaction logs, etc. It is a challenging task to identify the most relevant information, whose collection and analysis might bring the highest yield in the mid and long run. This PhD project will tackle the challenge by (1) identifying the relevant system processes (e.g. human-system interactions, intra-division processes, processes involving multiple bank divisions (system-system) and integrations with external systems), collecting and analysing their available data, (2) identifying sub-optimal internal processes, and (3) optimizing those processes wherever possible, thereby allowing financial systems to have measurable cost savings. The project will be implemented iteratively, and feedback will be given to the hosting financial institution (OTP) both about the quality and quantity of relevant information which they make accessible to this research. It will be possible to communicate the measurable impact in cost savings or other forms. The research will result in industry-grade solutions, applicable both in the target financial institution, but at other similar companies as well.
Innovation – *describe what is the intended solution and the advance w.r.t. the state-of-the-art*	Large financial institutions maintain complex information systems which consist of a myriad of different processes which are executed for and by different stakeholders. The day-by-day execution of those processes can be monitored, and potential bottlenecks can be identified. The intended solution will collect, clean and analyse the (big) data collected about real-life business processes and advance the state-of-the-art by leveraging the latest developments in the data science domain to identify pain points and bottlenecks which increase the workload and costs on one hand, and lower stakeholder satisfaction on

	the other hand. Once identified, solutions will be devised to address the weak links in the process ecosystem of financial systems. The research questions of the PhD will be formulated in line with financial institutions' need to optimize their internal processes, which is a true pain point, whose solution has potential to revolutionize how banks operate internally. A draft list of novel research questions which will be analysed during the PhD project follows: • **Process cost modelling**, i.e. novel methods to estimate the (financial) costs incurred by the financial institution when executing specific business processes, which range between simple bank account opening, implementing complex processes involving multiple departments in customer risk analysis or interactions with external entities (e.g. the national bank, tax authority). • **Sub-optimal process identification**, i.e. identify processes which contain excessive numbers of stakeholders or steps. Identify bottlenecks, long delays, participation of too many or too few parties, or process flows involving excessively high (financial) risk. • **Process optimization**, i.e. developing and proposing solutions which will allow financial systems to optimize their internal processes, thereby lowering costs and increasing stakeholder satisfaction.
Research focus/topics – *describe <u>how</u> you are going to solve the problem*	The most important research topics will be the following: 1) Data acquisition about the business processes which occur on various levels and between various actors of the (target) financial system. Initial analysis, industrial feedback and formulation of an initial set of hypotheses (e.g. process A involving 3 sub-systems costs more because of reasons X and Y). 2) Identify 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. 3) Create the first prototypes of models and automated data analysis and process optimization solutions. Test them on real-life business process datasets. 4) Improve the prototypes and deploy them in a production environment, or in an environment which is populated by real-time process data. 5) Evaluate the proper operation of the implemented solutions and make further developments based on empirical evidence, e.g. accuracy, satisfaction of key stakeholders in the financial institution, financial impact in the form of costs savings.
Deadlines/milestones (Gantt chart)	Review industrial best practices, current solutions and most recent research developments. Understand the business environment by identifying all relevant actors and stakeholders. Analyse the business processes and their underlying datasets. Specify the most important

	pain points (i.e. bottlenecks in processes) and create a list of hypotheses.
	Implement state-of-the-art baseline solutions developed by leading experts. Finalize the hypotheses. Receive feedback from OTP Bank about the baseline solutions and hypothesis. Update hypotheses based on the feedback received.
	Automate data collection, cleaning and structuring. Explorative data analysis, identification of sub-optimal business processes. Review the identified business process issues with the industrial partner.
	Develop and test the first prototypes of the predictive models. Extract data from the industrial partner's data warehouse for prototype verification. Present prototypes and receive feedback from OTP Bank.
	Verify and validate the solutions in a production environment. Optimize the prototype models and release their final, ready-to-be-industrialized versions. Transfer the results to OTP. Create a course on using the developed solutions in the re-engineering of sub-optimal processes.
	Summarize the results and complete the PhD thesis; deliver the course to OTP employees
Expected outcome – *describe the expected results of the PhD*	The results of the PhD will be incorporated into OTP Bank's and ELTE's joint EIT Innovation and Research projects in the Digital Finance domain. A toolbox capable to automatically collect, structure and clean process data, give model-based recommendations and predictions will be implemented. It will be transferred to OTP for its implementation in a financial system's production environment as a business support tool. The measurable outputs shall include: • Documentation of existing industrial solutions in the field of research. • Implementation of existing solutions utilized as baselines. • Business process optimization toolbox. • Process re-engineering course • A detailed analysis of the candidate's contribution and the impact on the bank's operational and business processes. • Published papers in journals and conferences.

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

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

Partnership/financial information

Action Line Leader	Antonio Garcia Hortal
Industrial partner	OTP Bank
Industry advisor – *name and short bio*	Illés Gozlán 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.
Academic/research partner	ELTE
Academic/research supervisor – *name and short bio*	Imre Lendák, PhD 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.
HEI granting the title	ELTE
DTC location	Budapest DTC
Geographical mobility plan	
No. of PhD positions	1
PhD duration	4
Co-funding percentages: - Industry	20
- Academia	30
- EIT Digital	50