Summer Intern 2018 positions at the department of Computer Science

We are recruiting a substantial number of summer interns to work on interesting research topics during the summer 2018. See full list of the topics below. You are supposed to express your preferences among the topics by specifying 1–5 preferred topics using the topic numbers available below.

Please, see the application instructions at: http://www.aalto.fi/en/about/careers/jobs/view/1719/
Deadline for applications is Feb 11, 2018 and the applications must be submitted through Aalto University’s recruitment system.

Topics
1. Topic: Systems security (multiple topics) ................................................................. 3
2. Topic: Optimization for large-scale linear solvers .................................................. 3
3. Topic: Hashing schemes for Deep and Large-scale Learning .................................. 3
4. Topic: Mobile Cloud Computing 3.0 ...................................................................... 3
5. Topic: Digitalizing Performance with Wearables and Software ............................. 3
6. Topic: Mobile Computing and the Internet of Things ............................................. 4
7. Topic: Analyzing and Predicting Human Health Status using Mobile Network Data .............................................................................................................. 4
8. Topic: Tools for Weak Memory Models ................................................................. 4
9. Topic: Big Data Pipeline for Pan-Genomics ........................................................... 4
10. Topic: Concurrent Programming Course Home Assignment Renewal ................. 5
11. Topic: Älykkääät elämäkerrat (Semantic Biographies) ........................................... 5
12. Topic: Computer vision .......................................................................................... 5
14. Topic: Algorithm theory and engineering for GPUs ............................................. 6
15. Topic: Approximate Bayesian Computation: inference on intractable models ........ 6
16. Topic: Interactive Learning for Personalized Medicine .......................................... 6
17. Topic: Privacy-preserving machine learning ......................................................... 7
19. Topic: Machine Learning for High-Dimensional and Relational Data, applied to Plant Breeding, Personalized Medicine or Bioinformatics .................................................. 7
21. Topic: Deep Gaussian processes .......................................................................... 8
23. Topic: Probabilistic programming with Stan ....................................................... 8
24. Topic: Topics on network science ......................................................................... 9
25. Topic: Statistical analysis of single cell data ............................................................ 9
26. Topic: Bayesian optimization for analysing single-cell mRNA time course data .................................................. 9
27. Topic: Bioinformatics methods for chromatin accessibility data analysis ............ 10
28. Topic: Deep kernels .............................................................................................. 10
29. Topic: Learning adaptive manifolds ................................................................. 10
30. Topic: Mathematical optimization techniques for machine learning............................... 11
31. Topic: Statistical machine learning methods for T1D biomarker discovery from multi-omics data sets ...... 11
32. Topic: Statistical machine learning methods to model human immune system ................. 11
33. Topic: Outdoor Cultural Heritage with Tangible Interaction ........................................ 12
34. Topic: Algorithms for the design of RNA nanostructures .................................................. 12
35. Topic: Open source tools for randomization and exploratory data analysis (2 positions) .......... 13
36. Topic: Hunt for silent signals in temporally structured data (1 or more positions) ............... 13
37. Topic: Peek into the black box (1 or more positions) ........................................................... 13
38. Topic: Constraint reasoning for finding subjectively interesting structure (1 or more positions) ... 14
39. Topic: Deep learning of particle physics experimental data (1 or more positions) ............... 14
40. Topic: Web of Things .............................................................................................. 15
41. Topic: Study on optimal imaging for 3D reconstruction ..................................................... 15
42. Topic: Declarative Programming and Optimization .......................................................... 15
43. Topic: Porting ASPTOOLS for the new ASPIF format ..................................................... 15
44. Topic: Automatic Deployment of Math Exercises ........................................................... 16
45. Topic: Points of Interest in Automated Course Materials ................................................ 16
46. Topic: Unified Programming Environments for Automated Assessment ................................ 16
47. Topic: Multi-modal and multi-lingual translation .............................................................. 17
48. Topic: Exercise problems and their automatic assessment in Basic Programming Y1 course ........ 17
49. Topic: Programming Assignment Developer for CS-A1140 Data Structures and Algorithms ........ 17
50. Topic: Games, movement AI, VR & AR ........................................................................ 18
51. Topic: Heavy Metal Machine Learning ............................................................................. 18
52. Topic: Content analysis and distribution in social media .................................................... 18
53. Topic: Interactive graph mining ...................................................................................... 18
54. Topic: Intelligent data analysis for network applications .................................................. 19
1. **Topic: Systems security (multiple topics)**

Our group ([https://ssg.aalto.fi/](https://ssg.aalto.fi/)) works on various aspects of systems security including platform security ("how to design and use hardware and OS security mechanisms to protect applications and services?"); security and machine learning ("what security and privacy concerns arise in machine-learning-based systems and how to address them?"); blockchains ("how to design scalable, efficient yet practical consensus mechanisms?", "what are compelling and novel applications for blockchain-based consensus?"). We also delve into other related security and privacy topics (e.g., "how to identify Internet trolls?") as well. To get an idea of the types of research topics available, take a look at [https://wiki.aalto.fi/display/sesy/Autumn+2017+-+Summer+2018](https://wiki.aalto.fi/display/sesy/Autumn+2017+-+Summer+2018)

Supervisor: N. Asokan
Email for more information: n.asokan@aalto.fi

2. **Topic: Optimization for large-scale linear solvers**

Large-scale linear classification is applicable in many supervised learning tasks involving such as text classification. The goal of the project would be to study large-scale linear solvers such as LibLinear ([https://www.csie.ntu.edu.tw/~cjlin/liblinear/](https://www.csie.ntu.edu.tw/~cjlin/liblinear/)) for large-scale text classification problems. In particular, the goal will be to study the sparsity inducing methods such as L1 regularization and shrinking heuristics used by Liblinear.

Supervisor: Rohit Babbar
Email for more information: rohit.babbar@aalto.fi

3. **Topic: Hashing schemes for Deep and Large-scale Learning**

With more data and overly complex deep models, the size of the models learnt is typically large. On the other hand, there is a need to compress the model size for its applicability and usage in power sensitive devices such as smart phones and IoT devices. The goal of the project would be to study the existing hashing techniques and build novel ones by utilizing the data and model sparsity for learning space and power efficient models. Reference: [https://arxiv.org/abs/1602.08194](https://arxiv.org/abs/1602.08194)

Supervisor: Rohit Babbar
Email for more information: rohit.babbar@aalto.fi

4. **Topic: Mobile Cloud Computing 3.0**

This summer internship is about developing the Mobile Cloud Computing course (CS-E4100) for Fall 2018. The student involved in this job is expected to: create online learning material using the A+ learning management system; employ a container-based framework for automated grading of course assignments; define and implement a proof-of-concept project related to mobile and cloud computing. Multiple positions available.

Supervisor: Mario Di Francesco
Email for more information: mario.di.francesco@aalto.fi

5. **Topic: Digitalizing Performance with Wearables and Software**

This summer internship is about building a software framework that bridges performance design and its realization on stage through embedded devices. The student involved in this job is expected to: use an indoor
location system to track a performer on stage; interface with DMX lighting control systems; build a simple user
to "configure" a performance. This job involves working with other students and/or professionals at
Aalto ARTS and the Theater Academy Helsinki.

Supervisor: Mario Di Francesco
Email for more information: mario.di.francesco@aalto.fi

6. Topic: Mobile Computing and the Internet of Things
This summer internship is related to current research on mobile computing and the Internet of Things. The
student involved in this job will conduct a research and development task in one of the following
topics: visible-light communications; edge and fog computing; architectures and protocols for the Internet of
Things (IoT); performance of IoT applications.

Supervisor: Mario Di Francesco
Email for more information: mario.di.francesco@aalto.fi

7. Topic: Analyzing and Predicting Human Health Status using Mobile Network Data
This summer internship is about exploring real-world data about mobile users to infer and anticipate their
health conditions. The student involved in this topic is expected to: predict physical and emotional health
symptoms (e.g., cold symptoms by using machine learning approaches); analyze correlation between physical
and emotional health status; analyze the spreading and evolution of disease epidemics.

Supervisor: Behrouz Jedari
Email for more information: behrouz.jedari@aalto.fi

8. Topic: Tools for Weak Memory Models
Weak memory models are a way to describe the programmer guarantees for multithreaded programs in Java,
C/C++, x86, Arm, Power, GPUs, and the Linux kernel. Our group has created novel tools to find concurrency
bugs in programs under weak memory models. Your task is to use and extend these tools to analyze low level
concurrency primitives found e.g., in the Linux kernel, trying to find bugs in the process. You will be joining a
research team that is actively working on the topic. Good programming skills are needed. Knowledge of
concurrent programming is a plus.

Supervisor: Assoc. Prof. Keijo Heljanko
Email for more information: keijo.heljanko@aalto.fi

9. Topic: Big Data Pipeline for Pan-Genomics
The research group has developed a parallel next generation sequencing data processing system, which uses
the Apache Spark framework to process human genomics next generation sequencing data in parallel, with
demonstrated scalability to 1280+ computing cores. In this project you will get involved with a new Spark
based pipeline that is implementing pan-genomics processing, which requires the genomes of several
individuals to be indexed into a single index, a feat that requires massively parallel processing power to be
accomplished on a large scale. You will join a research team working on this problem, programming parts of
the pipeline under the supervision of a PhD student working on the problem. You will need good programming
skills, with focus on Scala/Java, experience on Big Data platforms, especially Apache Spark is a plus.
10. Topic: Concurrent Programming Course Home Assignment Renewal
You will work on a new automated grading system for the course "CS-E4110 Concurrent Programming". This course teaches the students concurrent programming, and the autumn 2018 version is planned to have new automates home assignments for concurrent Scala programming exercises. You will create new home assignments along with unit testing suites that check students' code for grading purposes and provides useful feedback. You will gain experience in automated testing techniques for concurrent software as well as practical experience with Scala, Akka and Docker. Knowledge of Scala programming is required. Experience with concurrent programming is a plus.

Supervisor: Assoc. Prof. Keijo Heljanko
Email for more information: keijo.heljanko@aalto.fi

11. Topic: Älykkää elämäkerrat (Semantic Biographies)
[Finnish content is used in this position, and some understanding of Finnish is needed]

Supervisor: Prof. Eero Hyvönen
Email for more information: eero.hyvonen@aalto.fi, 050 3841618

12. Topic: Computer vision
Computer vision is a rapidly developing field that is at the forefront of recent advances in artificial intelligence. Our group has broad research interests within computer vision. We are pursuing problems both in geometric computer vision (including topics such as visual SLAM, visual-inertial odometry, image-based 3D modelling and localisation) and in semantic computer vision (including topics such as object detection and recognition, and deep learning). We are looking for students interested in both basic research and applications of computer vision. Students with good programming skills and strong background in mathematics are especially encouraged to apply. The precise topics of the research will be chosen together with the students to match their personal interests. Examples of our recent papers include: arxiv.org/abs/1708.00894, arxiv.org/abs/1707.09733 and arxiv.org/abs/1705.03386. For more papers and further information visit: https://users.aalto.fi/~kannalj1/

Supervisor: Juho Kannala
Email for more information: juho.kannala@aalto.fi
13. Topic: CS-A1120 Programming 2 courseware
Do you enjoy programming and mathematics? This position gives you the possibility to develop your skills further and participate in developing new programming exercises and other online infrastructure for the course "CS-A1120 Programming 2" at the Department of Computer Science. Having taken CS-A1120 yourself is an asset but not a prerequisite. Familiarity with the Scala programming language is essential.

Supervisor: Petteri Kaski
Email for more information: petteri.kaski@aalto.fi

14. Topic: Algorithm theory and engineering for GPUs
Modern high-performance computers are massively parallel: for example, an NVIDIA DGX-1 features 40960 32-bit scalar cores distributed across eight Tesla V100 accelerators linked with a fast interconnect. If you are interested in the mathematical design and engineering of high-performance algorithm implementations for cutting-edge hardware, such as a DGX-1, this is a position for you. Strong programming skills and mathematical maturity are essential. For an example of recent implementation engineering work, see https://doi.org/10.1137/1.9781611975055.16 and https://github.com/pkaski/six-subgraph

Supervisor: Petteri Kaski
Email for more information: petteri.kaski@aalto.fi

15. Topic: Approximate Bayesian Computation: inference on intractable models
We recently released ELFI, an engine for likelihood-free inference, with which it is possible to efficiently solve the hard task of fitting simulator-based models to data. At the moment, we are both continuing our basic research on the inference techniques, and applying the methods to new problems. We are looking for summer trainee to join our team to do both basic and applied research in probabilistic machine learning. Students having a strong background in mathematics and interest in modelling and inference are especially encouraged to apply. Additional information: http://elfi.readthedocs.io

Supervisor: Prof. Samuel Kaski
Email for more information: first.last@aalto.fi

16. Topic: Interactive Learning for Personalized Medicine
In this summer project, the goal is to improve as much as possible the prediction accuracy of drug effects, by interacting and obtaining feedback from an expert (doctor). Assuming a limited number of interactions with the expert, active learning techniques need to be used to identify and obtain the most relevant feedback for the expert. In other words, we develop new user interaction principles which combine machine learning with HCI for prior elicitation, that is, how can we extract the user’s prior knowledge to be included in probabilistic models with Bayesian inference. We are looking for a summer intern who has a strong background in machine learning and has basic programming skills. During the summer internship, you can contribute to model development and implementation of the models. Additional information: http://research.cs.aalto.fi/pml/

Supervisors: M.Sc. Iiris Sundin, Prof. Samuel Kaski
Emails for more information: first.last@aalto.fi
17. Topic: Privacy-preserving machine learning
We develop methods for learning from data given the constraint that privacy of the data needs to be preserved. This problem can be formulated in terms of a recent breakthrough called differential privacy, and we have introduced ways of doing the learning such that performance actually improves, in contrast to in alternative methods. A couple of “minor” problems still remain; come to solve them with us! Requirements: strong background in math, decent skills in programming, and/or a very steep gradient in the learning curve. Additional information: http://research.cs.aalto.fi/pml/

Supervisors: Dr. Teppo Niinimäki, Prof. Samuel Kaski
Emails for more information: first.last@aalto.fi

18. Topic: Machine learning for improved human-computer interaction
Adaptation to individual users’ needs and capabilities has potential to notably increase human capabilities in computerised activities. While conventional approaches require extensive trials and explicit user feedback, the proposed work focusses on model-based approaches where users’ cognitive and perceptual capabilities are inferred based on behavioural data, and optimum adaptations are chosen based on predicted user responses. In practice, we model how users complete interactive tasks and learn new behaviours, and use observed behavioural data to estimate model parameters that explain individual differences between users. The user interaction models considered in our work are based on methods like reinforcement learning, and model parameters are estimated with likelihood-free inference. We are looking for a summer trainee who wants to contribute to this new and promising research area that combines machine learning, information visualisation, and user interface design. Requirements: studies on machine learning or human-computer interaction, programming skills. Additional information: http://userinterfaces.aalto.fi, http://research.cs.aalto.fi/pml, http://www.helsinki.fi/bsg

Supervisors: Prof. Antti Oulasvirta, Prof. Samuel Kaski, Prof. Jukka Corander, Dr. Luana Micallef, Dr. Ulpu Remes
Emails for more information: antti.oulasvirta@aalto.fi

19. Topic: Machine Learning for High-Dimensional and Relational Data, applied to Plant Breeding, Personalized Medicine or Bioinformatics
In this summer project, you will work on developing a model of sparse matrix factorization and applying it to one or more biology-related problems in plant breeding, personalized medicine or more generally bioinformatics. Although the applications may sound diverse, the methodology behind them has the common idea of extracting low-rank approximation of given high-dimensional matrices/tensors/graphs. A spectrum of companies are collaboratively working with the applications, and company commitment is strong. This project requires good programming skills. Additional information: http://research.cs.aalto.fi/pml/

Supervisor: Prof. Hiroshi Mamitsuka
Emails for more information: first.last@aalto.fi

Interactive human-in-the-loop machine learning combines the skills and knowledge of humans with the computational and processing strengths of machines. We are developing new approaches and applications for interactive human-in-the-loop machine learning using probabilistic modelling methods, with the aim of
increasing the performance and efficiency of the systems and for improving the user experience. This project lies at the intersection of machine learning, human-computer interaction, and cognitive science. We are looking for a summer intern to work with us on this exciting and topical research problem. Requirements: knowledge of probabilistic methods in machine learning and programming skills. Additional information: 
http://research.cs.aalto.fi/pml/

Supervisors: Dr. Tomi Peltola, Prof. Samuel Kaski
Emails for more information: first.last@aalto.fi

21. Topic: Deep Gaussian processes
Deep neural networks have revolutionized machine learning as powerful predictors that can handle "google-scale" datasets. In a deep Gaussian process each layer is a Gaussian process, which is a powerful framework that combines interpretability of Bayesian learning and the power of deep architectures. In this summer project deep GP models will be developed and implemented for large-scale datasets or for instance to image recognition. Prior work on GP’s or image analysis is not required, but good knowledge of mathematics, programming and of Bayesian inference is very useful. Additional information: http://research.cs.aalto.fi/pml/

Supervisors: Dr. Markus Heinonen, Prof. Samuel Kaski
Emails for more information: first.last@aalto.fi

22. Topic: Probabilistic machine learning: post-hoc modeling of Bayesian models
An inherent property of probabilistic models is that complex models can be constructed by combining local submodels into a joint, global model. In this project, we develop methods for post-hoc inference in joint models, where the submodels themselves are given as individually computed posterior inferences. The framework has ties to distributed inference, transfer learning and information retrieval. Research tasks during the internship may include contributing to methodological development, devising scalable computational strategies, or investigating the theoretical properties of the framework. Requirements: strong background in math, decent skills in programming, and/or a very steep gradient in the learning curve.

Supervisors: Dr. Paul Blomstedt, Prof. Samuel Kaski
Emails for more information: first.last@aalto.fi

23. Topic: Probabilistic programming with Stan
Stan (http://mc-stan.org) is a probabilistic modeling language used by thousands of scientists, engineers, and other researchers for statistical modeling, data analysis, and prediction. It is being applied academically and commercially across fields as diverse as ecology, pharmacometrics, physics, political science, finance and econometrics, professional sports, real estate, publishing, recommender systems, and educational testing. In the summer project, you will take part in developing computational statistical methods related to model selection, variational inference or Gaussian processes in Stan. Strong background in mathematics and some or strong experience in programming is beneficial.

Supervisor: Aki Vehtari
Email for more information: first.last@aalto.fi
24. **Topic: Topics on network science**
There are several possible projects related to analysis of large scale networks on my research group. These topics usually involve data analysis, algorithm development or modelling work, and in the best case all of these. My research is mostly focused on applications of generalized network representations such as multilayer networks, temporal networks etc. We can select the exact topic based on the interests and competencies of the applicant.

Supervisor: Mikko Kivelä
Email for more information: mikko.kivela@aalto.fi (see also mkivela.com)

25. **Topic: Statistical analysis of single cell data**
During the past few years, the development of single cell sequencing techniques has revolutionized the systems biology field by allowing quantification of RNA and DNA content in individual cells. This new technology has important applications in revealing novel insights into basic molecular biology and various diseases, including cancer genomics.

We are looking for a self-motivated candidate to design and implement new statistical tools which could be integrated into our existing analysis framework for single cell data. More specifically, your task will be to (1) extend our existing hierarchical network inference method to be applicable also to large data sets and (2) tackle the related mathematical optimization problem.

You will work as an integral part of our multidisciplinary team and the method development work will be done together with team members specialized in mathematical and statistical modeling.

The project requires good knowledge of mathematics, statistics, and programming (e.g. Python, Matlab or R) as well as interest in molecular biology. For more information, see research group web page [http://research.cs.aalto.fi/csb/](http://research.cs.aalto.fi/csb/)

Supervisors: Dr. Jukka Intosalmi/Henrik Mannerström, Assist. Prof. Harri Lähdesmäki
Emails for more information: harri.lahdesmaki@aalto.fi, jukka.intosalmi@aalto.fi, henrik.mannerstrom@aalto.fi

26. **Topic: Bayesian optimization for analysing single-cell mRNA time course data**
The expression of single mRNA molecules can be monitored in real time at the single cell level by bioluminescence assays. To make the most of this data, models that incorporate the inherent stochasticity of the process are needed. Stochastic kinetic models (SKM) are continuous-time discrete state-space Markov models that are often used to model gene expression. However, parameter inference is exceedingly computationally expensive. For this model class, the likelihood function is intractable and other methods such as simulation has to be used.

Bayesian optimization is a probabilistic, state of the art method for finding the global optimum of a function. It can overcome problems such as noisy function evaluation and high cost of function evaluation. Applying Bayesian optimization to the parameter inference problem of SKM's is an interesting research topic that provides hands-on experience with state of the art optimization techniques and single cell data analysis.

The project requires knowledge of mathematics and programming (e.g. Matlab) as well as interest in molecular biology. For more information, see research group web page [http://research.cs.aalto.fi/csb/](http://research.cs.aalto.fi/csb/).
27. Topic: Bioinformatics methods for chromatin accessibility data analysis

In one human cell two meters of DNA are packed tightly within a five-micron nucleus while active DNA regions, such as promoters, enhancers and other regulatory elements need to be unpacked. These active regions are bound by transcription factors that regulate gene transcription. High-throughput measurement techniques, such as DNase-seq and ATAC-seq, combined with advanced computational methods provide us genome-wide information on DNA accessibility and transcription factor binding. On the other hand, R-loops, associated with open chromatin at promoters, are three-stranded nucleic acid structures. They are formed when a nascent RNA binds to template DNA leaving the non-template DNA single-stranded. R-loops are important regulators of gene expression but they can also promote DNA damage and are involved in neurological diseases and cancer.

Our aim in this project is to get a better understanding in R-loop formation by combining different types of sequencing data. Your task in this project is to familiarise yourself with the topic, especially high-throughput sequence data analysis, and participate in analyzing ATAC-seq data and developing novel statistical bioinformatics methods to predict R-loops and TF binding.

The project requires good knowledge of statistics and programming (e.g. R/Python/Matlab) as well as interest in molecular biology. For more information, see research group web page http://research.cs.aalto.fi/csb/

Supervisors: Sini Rautio, Prof. Harri Lähdesmäki
Emails for more information: harri.lahdesmaki@aalto.fi, sini.rautio@aalto.fi

-----------------------------------------------------------------------------------------------------------------------------------

28. Topic: Deep kernels

Deep neural networks have revolutionised machine learning as powerful predictors that can handle "google-scale" datasets, yet they are non-trivial to handle. More conventional and simpler Bayesian regression methods can be made equally powerful by equipping them with similarity functions inspired by neural network architectures. In this summer project we will explore covariances induced by neural architectures, and develop these into “deep”, rich kernels. Prior work on GP’s is not required, but good knowledge of math, programming and Bayesian inference is very useful. For more information, see research group web page http://research.cs.aalto.fi/csb/

Supervisors: Dr. Markus Heinonen, Prof. Harri Lähdesmäki
Emails for more information: markus.o.heinonen@aalto.fi, harri.lahdesmaki@aalto.fi

-----------------------------------------------------------------------------------------------------------------------------------

29. Topic: Learning adaptive manifolds

Underlying all machine learning is the concept of (metric) distance between vectorial data points. In many problems the distance between two data point changes over input space, or moves along a manifold. In this summer project we will learn Bayesian Gaussian process models that infer the model parameters, while simultaneously learning the underlying manifold or geometry of the data space. These models will be implemented and applied to problems of interest. Prior work on GP’s is not required, but good knowledge of math, programming and of Bayesian inference is very useful. For more information, see research group web page http://research.cs.aalto.fi/csb/
30. **Topic: Mathematical optimization techniques for machine learning**

During the last decades, machine learning has become an indispensable part of modern society. One of the most fundamental problems in the field is mathematical optimization, which can be defined as the data-driven numerical computation of parameters for a system designed to make decisions for unseen data. Numerous optimization algorithms have been proposed since the well-celebrated gradient descent, yet no recipe for selecting an appropriate optimization technique for a particular machine learning problem is present. Furthermore, each optimization method comes with its own assumptions and characteristics so that the methods could be classified in many respects: Being designed for convex or non-convex settings, fitting large-scale problems or not, being batch or stochastic, parallelizable or non-parallelizable, appropriate for high-dimensional problems or not, etc. In this project, student will examine variety of optimization techniques and apply them on machine learning research projects that are carried out in our group. The project requires good knowledge of mathematics, linear algebra, statistics, and programming (Matlab/Python/R). For more information, see research group web page [http://research.cs.aalto.fi/csb/](http://research.cs.aalto.fi/csb/).

Supervisors: Cagatay Yildiz, Assoc. Prof. Harri Lähdesmäki
Emails for more information: cagatay.yildiz@aalto.fi, harri.lahdesmaki@aalto.fi

31. **Topic: Statistical machine learning methods for T1D biomarker discovery from multi-omics data sets**

The recent advances in various ‘omics’ technologies and their increased affordability has enabled many study cohorts, such as DIABIMMUNE, to monitor various biological molecules in a high-throughput manner. The aim of the DIABIMMUNE project is to understand the development of type 1 diabetes (T1D) by following infants from Finland, Estonia and Russia from birth to 3-5 years of age. Various research groups across Finland and the world have been instrumental in collecting a diverse ‘omics’ datasets to study T1D, such as genomics, transcriptomics, epigenomics, proteomics, metagenomics, lipidomics, and metabolomics.

As highly informative as these datasets can be, the biological information that they capture is highly overlapping and no single ‘omics’ analysis can fully unravel the complexities of the fundamental biology of a disease. Therefore, in this study, you will apply advanced machine learning methods involving robust computational algorithms for integrating two or more ‘omics’ datasets in an attempt to understand T1D as one integrated system rather than as mere collections of different parts. For more information, see research group web page [http://research.cs.aalto.fi/csb/](http://research.cs.aalto.fi/csb/).

Supervisors: Juhi Somani, Prof. Harri Lähdesmäki
Emails for more information: juhi.somani@aalto.fi, harri.lahdesmaki@aalto.fi

32. **Topic: Statistical machine learning methods to model human immune system**

The immune response against pathogens is crucial for human health, but discrimination between self- and non-self-antigens can fail and result in an autoimmune disease. Each mature T cell has a unique T cell receptor (TCR) gene and this diversity defines how T cells discriminate between body's own antigens from pathogens, thus having a fundamental role in health and disease. Current technology now enables characterisation of individual's TCR repertoire. In this summer project student will explore and implement novel statistical machine learning methods to analyze individual's immune system properties and responses based
on high-throughput TCR sequencing data. The project requires good knowledge of machine learning, statistics and programming (e.g. R/Python/Matlab) as well as interest in molecular biology/biomedicine. For more information, see research group web page [http://research.cs.aalto.fi/csb/](http://research.cs.aalto.fi/csb/)

Supervisors: Emmi Jokinen, Dr. Markus Heinonen, Prof. Harri Lähdesmäki
Emails for more information: emmi.jokinen@aalto.fi, markus.o.heinonen@aalto.fi, harri.lahdesmaki@aalto.fi

33. Topic: Outdoor Cultural Heritage with Tangible Interaction

We've been working with Seurasaari Island to develop digital cultural heritage to help communicate the intangible heritage of the island. Last year we had an intern that started to build tangible boxes to allow triggered proximal interaction (e.g. using Bluetooth LE beacons). These might be used to ring an old rotary telephone that would deliver content or support interaction, or activate Olfactory emitters to augment a story of midsummer bonfires.

We are looking for a student with strong hacking (Arduino, raspberry PI) and programming skills (android, python) to develop these initial ideas further, also integrating them with an existing Android mobile app. Our goal is to evaluate their use in-situ at the island during the summer. This would suit a student with both strong skills and an interest in HCI. Additional information: [https://dl.acm.org/citation.cfm?id=3025803](https://dl.acm.org/citation.cfm?id=3025803) and [https://www.dropbox.com/s/qswqz6kp7vihgph/paper_v6_initial_submission_v1_new.pdf?dl=0](https://www.dropbox.com/s/qswqz6kp7vihgph/paper_v6_initial_submission_v1_new.pdf?dl=0)

Supervisor: David McGookin
Email for more information: david.mcgookin@aalto.fi

34. Topic: Algorithms for the design of RNA nanostructures


For several reasons, there is increasing interest in the DNA nanotechnology community to move from DNA to RNA as source material. This is however very challenging, because RNA has remarkably much richer and less well understood folding kinetics than DNA.

The topic of this internship project is to learn about the present combinatorial models of RNA folding and develop algorithmic methods for designing RNA sequences that fold into desired 2D or 3D shapes. Simulation studies will then be used to screen the proposed designs, towards a possibility of eventual validation by laboratory experiments.

The project requires familiarity with basic algorithm design techniques, facility with combinatorial thinking, and good programming skills. Previous knowledge of biomolecules is not necessary, although it is an asset. The work is performed in the context of research project "Algorithmic designs for biomolecular nanostructures (ALBION)”, funded by the Academy of Finland. For further information, please see the research group webpage at [http://research.cs.aalto.fi/no/](http://research.cs.aalto.fi/no/)

Supervisor: Prof. Pekka Orponen
35. Topic: Open source tools for randomization and exploratory data analysis (2 positions)
Visual exploration of high-dimensional datasets is a fundamental task in exploratory data analysis (EDA). We have developed a theoretical model for EDA, where patterns already identified and considered known by the user are input as knowledge to the system. The user is then shown views of the data where the user’s knowledge has been taken into account. Based on our recent work in EDA and randomization methods, the tasks in this project are twofold.

(i) Implement an open-source tool for exploratory data analysis. The tool should be web-based, cross-platform, and scale to large datasets.
(ii) Develop an open source library (e.g., in R, Python, JavaScript) implementing modern randomization methods for the use of data mining. Examples of such randomisation techniques include for instance maximum entropy models and different constrained randomisation schemes.

A natural division is for one person to implement the UI and for one to develop the algorithm library.

These tasks require good programming skills. Previous experience of open-source software development and knowledge of R, Python, JavaScript, HTML5, and frameworks such as React, Angular or Node.js are considered an advantage. The summer internship may continue to Master’s Thesis and even doctoral studies, depending on performance and funding.

Supervisors: Kai Puolamäki and/or Andreas Henelius and/or Emilia Oikarinen
Email(s) for more information: emilia.oikarinen@aalto.fi

36. Topic: Hunt for silent signals in temporally structured data (1 or more positions)
Assume that you have a stream of data where the task is to find silent signals, i.e., to spot changes in the signal as fast as possible, at the same time minimizing the chance of falsely declaring a change and the chance of missing a change. The task here is to study and develop methods based on randomization to accomplish this task in a principled manner. The data set addressed in the summer job is planned to include work accident data which is analyzed in collaboration with the Finnish Institute of Occupational Health.

We require a sufficient background in mathematics, probability and computer science. The summer internship may continue to Master’s Thesis and even doctoral studies, depending on performance and funding.

Supervisors: Kai Puolamäki and/or Andreas Henelius
Email for more information: andreas.henelius@aalto.fi

37. Topic: Peek into the black box (1 or more positions)
Modern artificial intelligence and machine learning techniques are very efficient and accurate on high-dimensional datasets. As a drawback, the methods are often black box models which are impossible for a human to interpret. This opens up several important and topical areas of research. For example:

(i) A classifier finds relevant structure from the data for class prediction. How can one study and use this structure for other purposes as well?
(ii) How can we interpret and explore the black box models? (A need for explaining the logic of model decisions is for instance mentioned by the new EU GDPR coming into effect in May 2018.)
(iii) Can the structures in the data exploited by the machine learning models be used to form better models?

Gaining deeper insight into black-box models is a central theme in our research group and the task(s) in this project are related to the above mentioned research questions.

This project requires a mathematical background, good understanding of machine learning and a good working knowledge of statistics. The summer internship may continue to Master’s Thesis and even doctoral studies, depending on performance and funding.

Supervisor: Andreas Henelius
Email for more information: andreas.henelius@aalto.fi

38. Topic: Constraint reasoning for finding subjectively interesting structure (1 or more positions)

Over the last decade there has been a growing interest in using generic constraint-based reasoning and optimization approaches (e.g., Boolean satisfiability, constraint programming, answer set programming, and linear programming) to solve data mining and machine learning tasks such as pattern mining, correlation clustering and learning optimal Bayesian networks. In this project the student will work in developing constraint reasoning based methods for data analysis that are optimised for use with humans and can be controlled by the humans, for instance, in terms of exact and efficient search for subjectively interesting structure in data using randomization.

A background in mathematics and interest in research is useful; prior knowledge of constraint-based reasoning methods (e.g. SAT, CP, ASP, LP), data analysis, machine learning, statistics and good programming skills are considered an advantage. The summer internship may continue to Master’s Thesis and even doctoral studies, depending on performance and funding.

Supervisor: Emilia Oikarinen
Email for more information: emilia.oikarinen@aalto.fi

39. Topic: Deep learning of particle physics experimental data (1 or more positions)

Your job is to analyze open data (http://opendata.cern.ch/) from a CERN particle physics experiment, and to study and develop data analysis methods to ultimately find previously unknown physics. The size of the full CMS Open Data set, part of which we use in this work, is over one petabyte. The prospective analysis methods include deep neural networks. This project is done in collaboration and in co-supervision with the group of Prof Mikko Voutilainen of the CERN CMS collaboration and University of Helsinki. There is a possibility the summer intern would do part of the assignment in CERN, Geneva, Switzerland, tentatively during June.

As prerequisites, we require knowledge of data analysis methods and a background in mathematics, probability, and programming. Studies in data sciences, computer science, and physics are considered an advantage. The summer internship may continue to Master’s Thesis and even doctoral studies, depending on performance and funding.

Supervisors: Kai Puolamäki and Mikko Voutilainen
40. Topic: Web of Things
We are developing a data sharing service for industrial internet. The service utilizes web technologies for device registration, data sharing contracts, and visualization. We are looking for one summer trainee to join the team. You can focus, for example, on time series databases, practical testing in various use cases, or data visualization.

Supervisor: Prof. Petri Vuorimaa
Emails for more information: petri.vuorimaa@aalto.fi

41. Topic: Study on optimal imaging for 3D reconstruction
The current internet is crowded with video content. For every WEB user, it is easily possible to develop own video content and upload them into the net. However, shooting videos in challenging environments or for special purposes where the best possible video content is required is not always straightforward.
In this project, we will explore how to enable best possible input video material for a special case where we develop 3D models of the indoor environment using video as the input information. Structure-from-Motion (SfM) is a technology that mitigates human visual system where 3D reconstruction in enabled when we film a certain object/scenery from different directions, likewise we see an object/scenery with our two eyes. Providing such material for a moving object or with moving camera exposes challenges, for example, how to ensure sharpness/focus of the video, or how to move the camera so that we maximize the quality of the 3D reconstruction. In this project, you will study and find out what are the parameters that impact on the quality of the 3D reconstruction, and you will develop rules and tools to optimize the 3D reconstruction system.

Supervisor: Prof. Antti Ylä-Jääski
Email for more information: antti.yla-jaaski@aalto.fi

42. Topic: Declarative Programming and Optimization
Declarative programming concentrates on specifying what is to be computed rather than how that computation should take place. A wide variety of problems arising in computer science and AI can be solved optimally using declarative specifications and suitable objective functions. Solutions can then be computed using existing general-purpose solvers. In this summer project, you are supposed to get acquainted with some constraint-based approach to declarative programming (to be agreed with the supervisor) in order to solve a particular application problem or problems. The actual tasks may involve the implementation of required tools for constraint processing, performing actual modeling with constraints, and benchmarking in order to compare the performance of different tool chains. Excellent programming skills are necessary and previous experience with constraint programming and AI is an asset.

Supervisor: Tomi Janhunen
Email for more information: tomi.janhunen@aalto.fi

43. Topic: Porting ASPTOOLS for the new ASPIF format
Over the years, the computational logic group at Aalto/CS has implemented a number of tools related with the Answer Set Programming (ASP) programming paradigm. Some of the tools are publicly available (see https://research.ics.aalto.fi/software/asp/) in the so-called ASPTOOLS collection. Quite recently, a new low-
level format (ASPIF) was introduced for ASP solvers, which makes older tools incompatible with the newest state-of-the-art solvers (such as clasp from the Potassco collection (see https://potassco.org/clasp/)). The goal of this summer project is to achieve the compatibility with the ASPIF format by developing the required C-library and by revising the basic data structures used by the ASPTOOL collection. Then the next objective is to port a number of tools on top of the new library and data structures. Good programming skills are required and previous experience with C/C++ and compiler technology is appreciated.

Supervisor: Tomi Janhunen
Email for more information: tomi.janhunen@aalto.fi

44. Topic: Automatic Deployment of Math Exercises
Math teachers are familiar and quite often also pleased with typesetting course materials and exercise questions in LaTeX. However, plain LaTeX representation is still quite far from formats that could be automatically deployed in web-based learning environments. The goal of this summer project is to make lower this barrier by introducing special LaTeX environments and macros that enable the teacher to express the aspects related with automation. The idea is that this extra information is extracted from the LaTeX representation of an exercise and used as the basis for automated installation, delivery for students, and automated evaluation. The implementation will be based on compiler technologies and macro processors. The approach is being developed for the purposes of some math courses organized at Aalto (acting as our customers). Good programming skills are required and previous experience with compiler compilers (flex, bison, or similar) is appreciated.

Supervisor: Tomi Janhunen
Email for more information: tomi.janhunen@aalto.fi

45. Topic: Points of Interest in Automated Course Materials
Web technologies enable the production of highly integrated and automated course materials that incorporate traditional lecture materials, exercises, programming projects, demos, and tools into one unified representation. The goal of this summer project is to develop extensions that help with the smooth presentation of the material to the students, e.g., during a lecture or a tutorial session. Roughly, the idea is that the teacher is able to navigate through points of interest that may also embed extra functionality that may be hidden from students. These extensions are implemented in the context of the A+ exercise framework (deployed by a number of computer science related courses) and Moodle. Good programming skills are required and previous experience with web technologies (including A+ and Moodle) and usability is an asset.

Supervisor: Tomi Janhunen
Email for more information: tomi.janhunen@aalto.fi

46. Topic: Unified Programming Environments for Automated Assessment
Many courses in computer science involve programming exercises that require particular programming languages, libraries, and tools to be used. From the students' perspective, this may require some installation work and, accordingly, the teacher may have to write very detailed instructions on how the installations are made and which versions of tools should be used. The goal of this summer project is to investigate possibilities for preparing unified programming environments (see, e.g., notebooks.csc.fi and www.abitti.fi) where all software and tools required by particular programming exercises/projects can be installed beforehand and
made uniformly available to students. Ideally, the same environment could be used in exams (where laptops would be used for programming) and exploited by automated graders as execution environments, using appropriate container technology. Good programming skills are required and previous experience in system administration and web technologies is appreciated.

Supervisor: Tomi Janhunen
Email for more information: tomi.janhunen@aalto.fi

47. Topic: Multi-modal and multi-lingual translation
Work together with our research group and collaborators at other universities to create neural sequence models that translate from the visual domain to human languages. Also, study how such image and video captioning models can be further used to facilitate translations between different languages. We will in particular develop and apply state-of-the-art deep learning techniques.

Supervisor: Jorma Laaksonen
Emails for more information: jorma.laaksonen@aalto.fi, mats.sjoberg@aalto.fi

48. Topic: Exercise problems and their automatic assessment in Basic Programming Y1 course
The task of the intern is to design new programming exercises for an elementary Python programming course, install them in A+ system, and implement their automatic assessment in A+ system. The course and the exercise problems are in Finnish. Thus, a good command of Finnish is necessary for this job.

Supervisor: Kerttu Pollari-Malmi
Email(s) for more information: kerttu.pollari-malmi@aalto.fi

49. Topic: Programming Assignment Developer for CS-A1140 Data Structures and Algorithms
The course was almost completely renewed in 2016 and now it is time to further improve its programming assignment system. Potential tasks include:
  i. developing new regular and extra "challenge" assignments [including assignment description, student code package and grading code]
  ii. porting selected assignments to C++ so that the students can (voluntarily) do some of the assignments in a high-performance low-level programming language [this includes porting both the code delivered to students as well as the grading code, and also testing and developing various C++ programming environments that could be offered to the students], and
  iii. developing a plugin system that allows students to (voluntarily) participate in a competition for the fastest solutions on some selected assignments in the course. The applicant should have good experience in both the Scala and C++ programming languages and be able to write clear problem descriptions and instructions in English; experience in Python is an additional merit.

Additional information: targeted for Aalto students only (might be continued as a TA position in Autumn 2018)

Supervisor: Tommi Junttila
Email for more information: tommi.junttila@aalto.fi
50. **Topic: Games, movement AI, VR & AR**
The intern will carry out a personal R&D project related to prof. Hämäläinen's group’s research, e.g.,
exergames, VR/AR sports, or movement AI for animation characters. The project will be tailored according to
the intern’s interests and background. We’re looking for a combination of CS & design skills, for example, a
student who could independently design and implement a VR simulation of soccer using a Unity & HTC Vive
with foot trackers.

Supervisor: Perttu Hämäläinen
Email for more information: perttu.hamalainen@aalto.fi

51. **Topic: Heavy Metal Machine Learning**

**Heavy metal and computer science**
Are you a computer scientist and like heavy metal music? Have you wondered why heavy metal and
computer science are usually not mentioned in the same sentence? Would you like to apply science to
heavy metal? In Modern Heavy Metal Music Studies (http://www.modern heavymetal.net/)
computer science and music come together in an academic discipline.

**Judging a band by their album covers**
In this Heavy Metal Music Studies project we are investigating the relation between album cover art
and musical genre. We are told to “not judge a book by its cover”, yet heavy metal bands develop a
distinct visual identity. Are these styles representative for different genres? If humans can recognise
different styles, are there common patterns? Can machine learning recognise these patterns and
identify or even predict musical styles from cover art?

**Machine learning of heavy metal album covers**
In this project, we will apply machine learning algorithms to album cover data sets. We will test if
unsupervised learning (e.g. clustering algorithms) can sort covers into meaningful subgroups. If time
permits, we will try supervised learning techniques (e.g. neural networks) to improve the genre
identification and prediction.

Computational Electronic Structure Theory (CEST) group in the Department of Applied Physics
Supervisors: Prof. Patrick Rinke, Dr. Amber Geurts and Dr. Heidi Henrickson
Emails for more information: firname.lastname@aalto.fi

52. **Topic: Content analysis and distribution in social media**
We are looking for motivated students to work on algorithms for content analysis and distribution in social
media. Topics of interest include analysis of social-media content, controversy and polarization in social
networks, echo chambers, dissemination of news in social media, opinion-formation models, and algorithms for
online content recommendation. Successful applicants are expected to have familiarity with graph mining,
machine learning, and/or combinatorial optimization. For additional information, see Data Mining group
website: http://research.cs.aalto.fi/dmg/

Supervisor: Aristides Gionis
Email for more information: aristides.gionis@aalto.fi

53. **Topic: Interactive graph mining**
We are looking for motivated students to work on algorithms for interactive graph mining, where humans (or in
general oracles) are part of the learning process, and the algorithmic objective is to develop efficient strategies
for better utilizing human resources and human knowledge. Successful applicants are expected to have
familiarity with graph mining, machine learning, and/or combinatorial optimization.
For additional information, see Data Mining group website: http://research.cs.aalto.fi/dmg/
54. Topic: Intelligent data analysis for network applications
We are looking for motivated students to work on machine-learning algorithms for anticipating network traffic in social-media applications and making intelligent choices to help improving the network, its services, and its performance. Successful applicants are expected to have familiarity with graph mining, machine learning, and/or combinatorial optimization. For additional information, see Data Mining group website: http://research.cs.aalto.fi/dmg/